No 189 (189) (2026)

The scientific heritage
(Budapest, Hungary)
The journal is registered and published in Hungary.
The journal publishes scientific studies, reports and reports about achievements in different scientific fields.
Journal is published in English, Hungarian, Polish, Russian, Ukrainian, German and French.
Articles are accepted each month.
Frequency: 24 issues per year.
Format - A4
ISSN 9215 — 0365

All articles are reviewed
Free access to the electronic version of journal
Edition of journal does not carry responsibility for the materials published in a journal.
Sending the article to the editorial the author confirms it's uniqueness and takes full responsibility for possible consequences for breaking copyright laws

Chief editor: Biro Krisztian
Managing editor: Khavash Bernat

- Singula Aleksandra - Professor, Department of Organization and Management at the University of Zagreb (Zagreb, Croatia)
- Bogdanov Dmitrij - Ph.D., candidate of pedagogical sciences, managing the laboratory (Kiev, Ukraine)
- Chukurov Valeriy - Doctor of Biological Sciences, Head of the Department of Biochemistry of the Faculty of Physics, Mathematics and Natural Sciences (Minsk, Republic of Belarus)
- Torok Dezso - Doctor of Chemistry, professor, Head of the Department of Organic Chemistry (Budapest, Hungary)
- Filipiak Pawel - doctor of political sciences, pro-rector on a management by a property complex and to the public relations (Gdansk, Poland)
- Flater Karl - Doctor of legal sciences, managing the department of theory and history of the state and legal (Koln, Germany)
- Bence Orban - Doctor of sociological sciences, professor of department of philosophy of religion and religious studies (Miskolc, Hungary)
- Feld Ella - Doctor of historical sciences, managing the department of historical informatics, scientific leader of Center of economic history historical faculty (Dresden, Germany)
- Owczarek Zbigniew - Doctor of philological sciences (Warsaw, Poland)
- Gál Jenő - MD, assistant professor of history of medicine and the social sciences and humanities (Budapest, Hungary)
- Borbély Kinga - Ph.D, Professor, Department of Philosophy and History (Kosice, Slovakia)
- Eberhardt Mona - Doctor of Psychology, Professor, Chair of General Psychology and Pedagogy (Munich, Germany)
- Kramarchuk Vyacheslav - Doctor of Pharmacy, Department of Clinical Pharmacy and Clinical Pharmacology (Vinnytsia, Ukraine)

«The scientific heritage»
Editorial board address: Budapest, Kossuth Lajos utca 84,1204
E-mail: public@tsh-journal.com
Web: www.tsh-journal.com

CONTENT

BIOLOGICAL SCIENCES

Arabova G., Jabrailzade S.
DIVERSITY OF PHYTOPATHOGENIC FUNGI ASSOCIATED WITH CULTIVATED *PISTACIA VERA* L. IN THE GREATER CAUCASUS.....3

CHEMISTRY SCIENCES

Klimko Yu., Koshchii I.
SYNTHESIS OF NEW ADAMANTYL-CONTAINING PYRAZOLONE CHALCONE.....7

HISTORICAL AND ARCHEOLOGICAL SCIENCES

Aliyev F.
ON THE ISSUE OF THE ECONOMIC ACTIVITY OF THE SHIRVANSHAH DYNASTY IN THE HISTORICAL ASPECT DURING THE EARLY MIDDLE AGES.....11

Aliyev A., Kerimov Z., Yevseyev R., Ibragimov A.
FEMALE FACES FROM MEDIEVAL NAKHCHIVAN.....15

MEDICAL SCIENCES

Shkiryak A., Khanenko S.
THE POTENTIAL ROLE OF SERUM VITAMIN D LEVEL IN MIGRAINE HEADACHE: A CASE–CONTROL STUDY....20

Shkiryak A., Khanenko S.
PROBABILITY OF VITAMIN D DEFICIENCY BY BODY WEIGHT AND RACE/ETHNICITY.....25

PHYSICS AND MATHEMATICS

Liu Zhaoxia
PRACTICAL APPLICATION ANALYSIS OF DERIVATIVES IN OPTIMIZATION PROBLEMS.....30

PSYCHOLOGICAL SCIENCES

Akanov T., Dupliakin E.
GESTALT THERAPY IN WORKING WITH EXISTENTIAL MIDLIFE CRISES: A NARRATIVE REVIEW.....32

TECHNICAL SCIENCES

Kotyk M., Kukuruza V.
PHYSICS-CONSTRAINED SURROGATE OPTIMIZATION OF MEMS PIEZORESISTIVE PRESSURE SENSOR WITH CO-DESIGNED DIGITAL FILTERING CHAIN FOR HIGH-PERFORMANCE BIOMEDICAL APPLICATIONS 36

Kotyk M., Kukuruza V.
A FULLY OPEN-SOURCE WORKFLOW FOR THE DESIGN AND VALIDATION OF MEMS PIEZORESISTIVE PRESSURE SENSORS: ANALYTICAL THEORY, PURE-PYTHON PLATE FEM, AND 3-D CONTINUUM SIMULATION.....51

Perelekhov I.
DESIGNING FAULT-TOLERANT E-COMMERCE SYSTEMS: ARCHITECTURAL PATTERNS AND PRACTICAL SOLUTIONS 58

BIOLOGICAL SCIENCES

DIVERSITY OF PHYTOPATHOGENIC FUNGI ASSOCIATED WITH CULTIVATED *PISTACIA VERA* L. IN THE GREATER CAUCASUS

Arabova G.
Azerbaijan State Pedagogical University, Ph.D in Biology
Jabrailzade S.
Azerbaijan State Pedagogical University, Accociate professor
DOI: 10.5281/zenodo.21617793

Abstract

The study investigates the species diversity, ecological traits, and trophic strategies of phytopathogenic fungi associated with *Pistacia vera* L. orchards in the Greater Caucasus region. The complex vertical zonation, along with climatic and soil diversity, creates favorable conditions for rich biodiversity, particularly of fungi. Increasing altitude affects temperature, precipitation, soil formation, and organic matter accumulation, influencing pathogen development. Although pistachio is a perennial plant adapted to semi-arid conditions, it is susceptible to leaf necrosis, fruit rot, vascular wilting, and stem diseases, reducing photosynthesis and productivity. Approximately 100 samples yielded 63 pure cultures, representing 38 fungal species, of which 20 were phytopathogenic; 13 belonged to Ascomycota and 7 to Basidiomycota, indicating ascomycete predominance. The results highlight a rich but potentially risky fungal complex in pistachio orchards, providing a basis for early disease diagnosis, integrated management strategies, and sustainable cultivation. Future studies should focus on molecular identification and the effects of climate variability on pathogen diversity.

Keywords: agroecosystems, Greater Caucasus, Pictacia vera L., mycobiota, orchard diseases.

İntroduction. The Greater Caucasus is located in the northern part of Azerbaijan and forms one of the most important mountain systems of the country. This extensive territory creates a wide climatic range according to the law of altitudinal zonation. Thus, from the lower mountainous zones to the high mountain areas, different climatic variations are observed across regions, ranging from a dry climate with hot summers to a cold climate with severe winters in different seasons.

As altitude increases, the decrease in temperature and the increase in precipitation directly affect soil formation processes and the structure of vegetation cover. While mountain forest soils dominate in the middle mountain zones, peat and sod mountain-meadow soils are widely distributed in the subalpine belt. High humidity and rich organic matter reserves create favorable ecological conditions for the development of microorganisms, including biotrophic fungi. The climatic diversity of the region is one of the key factors determining ecosystem stability and the distribution dynamics of phytopathogenic fungi.

Pistachio (*Pistacia vera* L.) is a long-lived and economically important nut-bearing plant belonging to the family Anacardiaceae. It is highly adapted to semi-arid and arid climatic conditions and, due to its strong root system, can absorb water from deep soil layers and tolerate temperature fluctuations. Pistachio fruits are rich in protein, unsaturated fatty acids, vitamins, and antioxidants, making them strategically important for both food consumption and agricultural export.

The main pistachio-producing regions in the world include the Mediterranean basin, the Middle East, Central Asia, and the United States. In recent years, the Greater Caucasus region of Azerbaijan has also been considered a promising cultivation zone due to its favorable climatic and soil conditions (Nouri et al., 2020; Rhouma, 2024).

Nevertheless, pistachio is susceptible to a number of biotic stress factors, particularly phytopathogenic fungi. These pathogens can cause leaf necrosis, fruit rot, vascular wilting, and branch dieback, leading to reduced photosynthesis and decreased productivity. Biotrophic fungi, which obtain nutrients exclusively from living plant tissues and establish long-term interactions with their hosts, are of particular interest. The study of these pathogens is important for the development of sustainable disease management strategies, the selection of resistant cultivars, and the maintenance of orchard productivity (Crous et al., 2013; López-Moral et al., 2022).

The Greater Caucasus region is characterized by complex altitudinal zonation, which leads to significant variations in climate, soil, and vegetation cover. In the lower and middle mountain zones, a moderately subtropical climate and humus-rich soils predominate, while in the high mountain areas temperatures are lower, precipitation levels are higher, and the soils are characterized by poorly developed alpine soils (Botany.az, 2023). These conditions directly influence the growth of pistachio trees and the diversity of fungi associated with them. High humidity and rich organic matter reserves create a favorable microclimate for the development of biotrophic fungi.

Studies conducted in neighboring countries show that in pistachio-producing regions such as Turkey and Iran, *Septoria* spp., *Pseudocercospora pistacina*, *Verticillium dahliae*, *Alternaria alternata*, and rust fungi are among the main biotrophic and hemibiotrophic pathogens affecting pistachio orchards (Drais et al., 2023; Gusella et al., 2022). However, information about the ecological diversity and distribution potential of these fungi in the Greater Caucasus region remains limited.

The aim of this study is to analyze the diversity of pathogenic fungi associated with pistachio plants cultivated in the Greater Caucasus region, to highlight their

ecological significance, and to discuss their implications for plant health and sustainable orchard management. The integration of the region's ecological characteristics with available data on fungal diversity provides a basis for future research focused on early pathogen detection, integrated disease management, cultivar selection, and agroecological optimization.

Material and methods. Samples were collected from cultivated pistachio trees growing in the economic–geographical regions located within the Greater Caucasus of Azerbaijan, based on the planned route and permanent plot selection methods widely used in mycological studies (Томашевич, М. А., 2015). During sampling, specimens were taken from the aboveground vegetative organs (stems and leaves) as well as generative organs (flowers and fruits) of the plants. The collected samples were documented on-site, labeled, and packaged for further laboratory analyses.

Under laboratory conditions, the analysis of the fungal biota of the collected samples began with the isolation of pure fungal cultures. The purity of the cultures was monitored using a microscope. To obtain pure cultures of fungi from the collected samples, several standard nutrient media were used, including agarized malt extract medium (MEA), potato agar (PA), rice agar (RA), agarized Czapek medium (CZA), and Sabouraud agar (SA). After inoculation onto Petri dishes containing the nutrient media, the samples were placed in a thermostat at a temperature of 25–28 °C for a period of 3–12 days and maintained until fungal colonies developed.

The species composition of the isolated pure cultures was determined using both macroscopic and microscopic methods. In this process, identification keys and atlases based on the cultural–morphological and physiological characteristics of fungi were used. The systematic affiliation and nomenclature of the fungi were verified according to the classification system provided on the official website of the International Mycological Association (https://www.mycology.org), as well as the data available in Index Fungorum (http://www.indexfungorum.org).

The identification of the diseases caused by the fungi and their respective causal agents was also carried out using classical phytopathological methods with the help of established identification manuals.

Results and conclusion. The mycological analysis of samples collected from cultivated pistachio orchards in the Greater Caucasus region resulted in the isolation of 63 pure fungal cultures obtained from approximately 100 plant samples. The samples were collected from different vegetative and generative organs of the plant, including leaves, stems, flowers, and fruits. As a result of morphological and microscopic identification, 38 fungal species belonging to different taxonomic groups were determined.

Among the identified fungi, 20 species were classified as phytopathogenic fungi associated with the cultivated pistachio tree (Pistacia vera). Taxonomic analysis showed that 13 species belong to the division Ascomycota, while 7 species belong to Basidiomycota, indicating the dominance of ascomycetous fungi within the phytopathogenic complex.

The phytopathogenic fungi identified in the study include species such as *Alternaria alternata, Alternaria tenuissima, Colletotrichum gloeosporioides, Cytospora carposperma, Fusarium dimerum, Fusarium moniliforme, Monilia fructigena, Phomopsis prunorum, Septoria pistaciae,* and *Podosphaera pistacia.* These fungi are known to cause different plant diseases affecting pistachio orchards in various climatic regions.

The pathological manifestations observed in infected plants included leaf spot, powdery mildew, septoria leaf spot, anthracnose, stem lesions, fruit rot, and wood decay. Among these diseases, leaf spot symptoms were the most frequently observed, indicating that foliar pathogens represent a dominant component of the phytopathogenic complex in the studied region.

Fungal pathogens identified during the research also differ in their substrate specificity. Some species demonstrate high host specificity and are mainly associated with pistachio trees. For example, *Podosphaera pistacia, Monilia pistaciae,* and *Septoria pistaciae* are considered host-specific fungi that primarily infect pistachio plants. In contrast, other species such as *Alternaria alternata* and *Fusarium spp.* are known as polyphagous pathogens, capable of infecting a wide range of host plants.

From an ecological perspective, the detected fungal pathogens were classified according to their trophic strategies, which include biotrophic, hemibiotrophic, and necrotrophic lifestyles. Biotrophic fungi obtain nutrients from living host tissues and maintain a prolonged association with the plant. In this study, species such as *Podosphaera pistacia, Fomes fomentarius, and Melampsora fagi* were categorized as biotrophic fungi. Hemibiotrophic fungi, which initially establish a biotrophic relationship and later switch to a necrotrophic phase, include *Colletotrichum gloeosporioides, Cytospora carposperma, Phomopsis prunorum, Septoria pistaciae, and Sphaerulina pruni.* These pathogens are capable of causing severe disease symptoms due to their dual trophic strategy. Necrotrophic fungi, which obtain nutrients from dead plant tissues, were also widely represented among the isolated pathogens. Species such as *Alternaria alternata, Alternaria tenuissima, Fusarium dimerum, Fusarium moniliforme, Monilia fructigena, and Ascochyta fagi* belong to this ecological group.

Phytopathogenic fungi associated with cultivated pistachio (Pistacia vera) in the Greater Caucasus region

Fungal species	Division	Disease / pathology	Trophic strategy
Alternaria alternata	Ascomycota	Leaf spot	Necrotroph
Alternaria tenuissima	Ascomycota	Leaf spot	Necrotroph
Ascochyta fagi	Ascomycota	Ascochyta leaf blight	Necrotroph
Colletotrichum gloeosporioides	Ascomycota	Anthracnose	Hemibiotroph
Cytospora carposperma	Ascomycota	Stem canker	Hemibiotroph
Fusarium dimerum	Ascomycota	Root rot / wilt	Necrotroph
Fusarium moniliforme	Ascomycota	Vascular wilt	Necrotroph
Monilia fructigena	Ascomycota	Fruit rot	Necrotroph
Monilia pistaciae	Ascomycota	Pistachio fruit rot	Hemibiotroph
Phomopsis prunorum	Ascomycota	Shoot blight	Hemibiotroph
Podosphaera Pistacia	Ascomycota	Powdery mildew	Biotroph
Septoria pistaciae	Ascomycota	Septoria leaf spot	Hemibiotroph
Sphaerulina pruni	Ascomycota	Leaf spot	Hemibiotroph
Fomes fomentarius	Basidiomycota	White wood rot	Biotroph
Fomitopsis pinicola	Basidiomycota	Brown wood rot	Saprotrophic/ necrotrophic
Laetiporus sulphureus	Basidiomycota	Wood decay	Necrotroph
Melampsora fagi	Basidiomycota	Rust disease	Biotroph
Phellinus igniarius	Basidiomycota	White trunk rot	Necrotroph
Schizophyllium commune	Basidiomycota	Wood decay	Necrotroph
Trametes versicolor	Basidiomycota	White rot	Necrotroph

An important ecological distinction among the detected fungi relates to the concept of ecological and physiological biotrophy, as described in previous mycological studies (Crous et al., 2013). For instance, macrofungi such as *Fomes fomentarius* demonstrate ecological biotrophy and can be successfully cultivated on artificial nutrient media. In contrast, physiologically biotrophic fungi, such as *Melampsora fagi,* cannot be easily cultured under standard laboratory conditions because they require living host tissue for development.

The diversity of phytopathogenic fungi associated with pistachio orchards in the Greater Caucasus reflects the complex ecological conditions of the region, including variations in altitude, humidity, temperature, and soil composition. The vertical zonation characteristic of the Greater Caucasus creates diverse ecological niches that support the development and persistence of different fungal species.

The coexistence of biotrophic, hemibiotrophic, and necrotrophic fungi within pistachio orchards indicates that plant diseases in this region develop through multiple infection mechanisms. Such diversity of pathogenic strategies increases the complexity of disease management and highlights the importance of integrated plant protection approaches.

Overall, the results of this study demonstrate that pistachio plantations in the Greater Caucasus region host a diverse phytopathogenic fungal community. The identified pathogens represent a potential risk factor for pistachio cultivation, although the current phytosanitary situation can be considered moderately stable and manageable with proper monitoring and control strategies.

These findings provide an important scientific basis for the early diagnosis of fungal diseases, development of integrated disease management strategies, and selection of resistant pistachio varieties. Future studies should focus on molecular identification methods, population genetics of pathogens, and the impact of climate change on fungal diversity, which would further contribute to the sustainable development of pistachio cultivation in the region.

References

1. Bakshaliyeva, K.F. (2024) General characteristics of some fruit plants included in the flora of Azerbaijan and their mycobiota/ K.F. Bakshaliyeva, G.Q. Arabova, E.O. Iskandar [et al.].//Advanced Studies in Biology (Bulgaria), vol. 16, № 1, -p.35-43.

2. Barnett, H. L., & Hunter, B. B. (1998). *Illustrated Genera of Imperfect Fungi* (4th ed.). APS Press.

3. Booth, C. (1971). *The Genus Fusarium*. Commonwealth Mycological Institute

4. Botany.az. (2023). *Böyük Qafqazın flora və torpaq xüsusiyyətləri*. Azərbaycan Dövlət Pedaqoji Universiteti, Bakı.

5. Crous, P. W., P. W., Gams, W., Stalpers, J. A., Robert, V., & Stegehuis, G. (2013). MycoBank: an online initiative to launch mycology into the 21st century. *Studies in Mycology*, 50, 19–22.

6. Drais, B., B., et al. (2023). Fungal diseases of pistachio in Mediterranean regions: diversity and epidemiology. *Journal of Plant Pathology*, 105(2), 345–358.

7. Gusella, G., G., et al. (2022). Emerging fungal pathogens associated with pistachio decline in the Middle East. *Plant Disease*, 106(4), 1120–1132.

8. Index Fungorum. (2023). *Global fungal names database*. Retrieved from http://www.indexfungorum.org

9. International Mycological Association (IMA). (2023). *Official fungal nomenclature database*. Retrieved from https://www.mycology.org

10. Kirk, P. M., Cannon, P. F., Minter, D. W., & Stalpers, J. A. (2008). *Dictionary of the Fungi* (10th ed.). CABI Publishing.

11. López-Moral, A., A., Raya-Ortega, M. C., & Trapero, A. (2022). Biology and control of pistachio fungal pathogens: current status and future perspectives. *Crop Protection*, 158, 106012.

12. Nouri, A., A., et al. (2020). Climate suitability and sustainable pistachio cultivation in semi-arid regions. *Agricultural Systems*, 178, 102742.

13. Rhouma, A., A. (2024). Global pistachio production and fungal disease challenges under climate change. *Horticultural Science*, 59(1), 15–28.

14. Seifert, K., Morgan-Jones, G., Gams, W., & Kendrick, B. (2011). *The Genera of Hyphomycetes*. CBS-KNAW Fungal Biodiversity Centre.

15. Бондарцева, М.А. (1998). Определитель грибов России. Порядок афиллофоровые/М.А. Бондарцева. -СбП. : Наука, в. 2, -391с.

16. Гарибова, Л.В. (2009). Популярный атлас-определитель грибов/Л.В.Гарибова. -М.: Просвещение / Дрофа, -352с.

17. Томашевич, М.А. (2015). Формирование патокомплексов растениц при интродукции в Сибири:/диссертации д.б.н./-Новосибириск, с.462

CHEMISTRY SCIENCES

SYNTHESIS OF NEW ADAMANTYL-CONTAINING PYRAZOLONE CHALCONE

Klimko Yu.
PhD in Chemistry, Senior Lecturer
Department of Organic Chemistry and Technology of Organic Substances
National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute"
Koshchii I.
PhD in Chemistry, Associate Professor
Department of Organic Chemistry and Technology of Organic Substances
National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute"
DOI: 10.5281/zenodo.21617810

Abstract

Heterocyclic compounds play an important role in drug discovery and development. In this study, a new pyrazolone chalcone (3) was synthesized by reacting azopyrazolone (2) with 4-(1-adamantyl)benzaldehyde in an alkaline medium. Numerous methods were used to characterize the target product, including elemental analysis, ^{1}H NMR spectroscopy, ^{13}C NMR spectroscopy, and IR spectroscopy. Compound (3) may serve as a potential anticancer agent.

Keywords: pyrazolone chalcone, adamantyl, anticancer agent, ^{1}H and ^{13}C NMR spectra.

Introduction

Cancer is a global public health problem, and its early diagnosis is of great importance [1]. Currently, the main strategies for cancer treatment are chemotherapy, surgery, and targeted therapy. Targeted therapy has become a standard treatment for cancer patients and has significantly improved the efficacy of anticancer therapy [2]. According to previous research, the regulation of organ growth and tumor suppression depend on the control of cell division and proliferation by the targeted Hippo signaling pathway [3]. Due to its significant impact on disease prognosis and resistance to chemotherapeutic drugs, the Hippo pathway is a desirable therapeutic target [4]. Pyrazoles have attracted considerable attention due to their effective pharmaceutical properties, such as antibiotic, antifungal, antibacterial, pesticidal, antioxidant, anti-inflammatory, antiviral, and antitumor activities [5-11]. Chalcones are α,β-unsaturated ketones with a broad spectrum of pharmacological activity, making them interesting synthons and bioactive structures with great therapeutic potential. They are well known to possess significant biological effects, such as antibacterial, antioxidant, and anticancer activity [12–15]. Chalcones facilitate the synthesis of new heterocyclic compounds, including pyrimidinones, азепины, пиридины и пираны, которые обладают сильной биологической активностью [16,17]. Существует множество сообщений о продаваемых противораковых препаратах, содержащих пиразолон и халконовое ядро, таких как метохалкон, софалкон, бутеин, метамизол и пропифеназон [18–22] как показано на рис. 1.

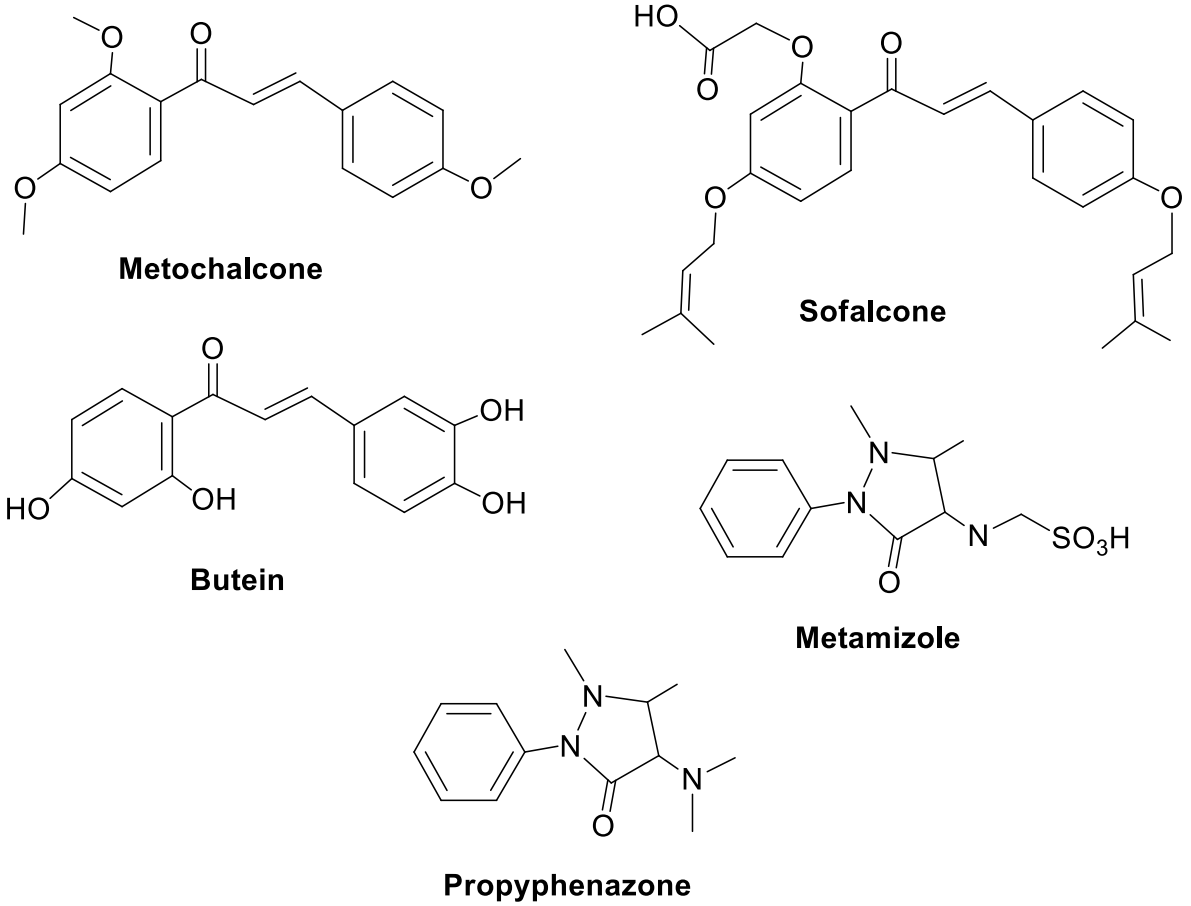
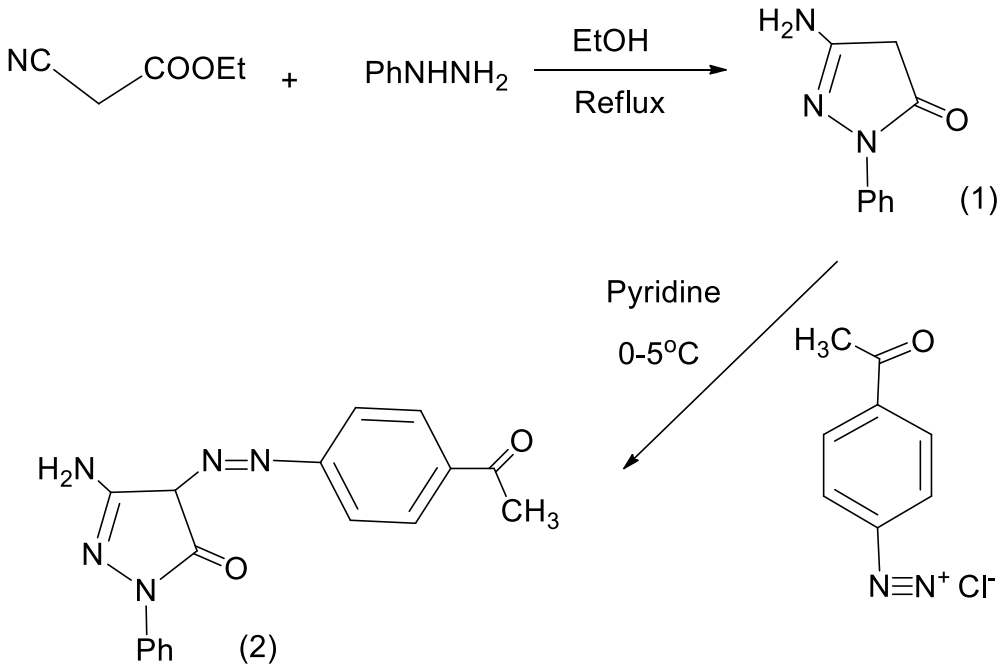


Fig. 1 Marketed pyrazolone- and chalcone-based drugs.

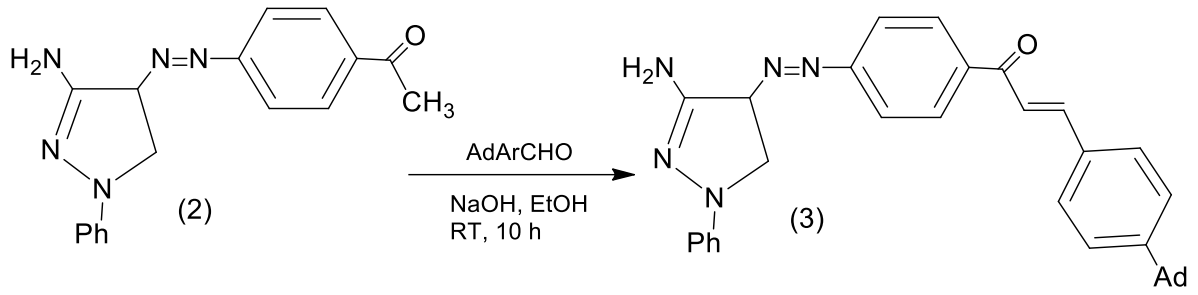
The aim of our study was to synthesize a new pyrazolone chalcone derivative with increased lipophilicity due to the presence of an adamantyl substituent.

Results and discussion

Scheme 1 describes the reaction of ethyl cyanoacetate with phenylhydrazine, which leads to the formation of compound (1), as described previously by Weissberger [23]. Further modification of compound 1 with p-acetylphenyldiazonium chloride led to the formation of azopyrazolone (2) in 74% yield (Scheme 1). The structure of compound (3) was established based on elemental analysis and various spectral data. Fourier transform infrared spectroscopy (FT-IR) was used for this purpose.



Scheme 1 Synthesis route for compound 2.



Scheme 2 Synthesis route for compound (3).

The different absorption bands at 1600 cm^{-1} and 1540 cm^{-1} are due to the C=O and N≡N groups, respectively. Modification of the azo compound (2) with another aromatic aldehyde in an alkaline medium led to the formation of pyrazolone chalcone (3) (Scheme 2) [24]. IR spectroscopic analysis of chalcones (3) showed absorption bands at 1610–1670 cm^{-1}, 1520–1600 cm^{-1} and 1500–1560 cm^{-1} for the C=O, C=N and N=N groups. The ^1H NMR spectrum of compound (3) revealed two new signals resonating at δ 7.54 and 7.68 ppm, which were assigned to two olefinic protons.

Experimental section

Materials and instrumentation

The starting compounds were purchased from Aldrich and used as received. All solvents were of analytical grade and distilled before use. IR spectra were recorded in KBr on a Bruker-ALPHA-II spectrophotometer. ^1H and ^{13}C NMR spectra were obtained on a Bruker 400 MHz NMR spectrometer. Chemical shift values are reported in parts per million (ppm) using TMS as the internal standard and DMSO-d6/$CDCl_3$ as solvents.

Synthesis of 3-amino-1-phenyl-1H-pyrazol-5(4H)-one (1).

Weissberger lab was previously preparing compound (1) [23].

Synthesis of 4-((4-acetylphenyl)diazenyl)-3-amino-1-phenyl-1H-pyrazol-5(4H)-one (2).

A sodium nitrite solution (12.7 mmol) was added slowly to pamino acetophenone (13.7 mmol) in concentrated HCl. The resulting diazonium chloride was added to a solution of compound 1 (8.5 mmol) in pyridine. Then, the product was filtered and dried. Orange powder, yield 74%; mp 89–91 °C; 1H NMR (400 MHz, DMSO-d6) d (ppm): 9.14 (s, 2H, NH2), 7.08–8.11 (m, 9H, Ar–H), 2.59 (s, 1H, CH–pyrazolone), 2.20 (s, 3H, CH3); IR (KBr) n: 1600 (CO), 1540 (N=N). Anal. calcd for $C_{17}H_{15}N_5O_2$ (321.33): C, 63.54%; H, 4.71%; N, 21.79%. Found: C, 63.34%; H, 4.57%; N, 21.67%.

5-Amino-4-((E)-(4-cinnamoylphenyl)diazenyl)-2-(4-(1-adamantyl)phenyl)-2,4-dihydro-3H-pyrazol-3-one (3).

An ethanolic solution of compound 2 (10.0 mmol), 4-(1-adamantyl)benz aldehyde (10.0 mmol), and sodium hydroxide (20.0 mmol) was stirred at 25 °C for 10 h, put into freezing water, filtered, and then dried. Orange powder, yield 76%. M.p. 248–250 °C; ^1H NMR (400 MHz, DMSO-d6) δ(ppm): 10.55 (s, 2H, NH_2), 7.68 (d, 1H, =CH), 7.54 (d, 1H, COCH=), 6.54–8.40 (m,14H, Ar–H), 2.60 (s, 1H, CH–pyrazolone), 1.92, 1.68, 1.48 (m, 15H Ad). ^{13}C NMR (101 MHz, DMSO-d6) d (ppm): 171.72 (C=O chalcone), 168.04 (C=O pyrazolone), 146.95 (C=N pyrazolone), 113.20–133.22 (Ar–C) 127.60, 135.25 (C–olefinic), 74.43 (CH–pyrazolone), 45.3, 37.5, 31.6, 29.0 (Ad-C). IR (KBr) ν: 3405 (NH_2), 1650 (C=O), 1597 (C=N), 1550 (N=N). Anal. calcd for $C_{34}H_{33}N_5O_2$ (543.15): C, 75.14%; H, 6.08%; N, 12.89%. Found: C, 75.25%; H, 6.06%; N, 12.82%.

Conclusion

1-Adamantyl-containing pyrazolone chalcone (3) was synthesized and characterized using various spectral data.

References

1. P. Das, G. Sayan, A. Tarun, M. Pritiprasanna, G. Sabyasachi, C. Sumita and G. Subhashis, Mater. Chem. Phys., 2019, 237, 121860.

2. W. Cao, H. D. Chen, Y. W. Yu, N. Li and W. Q. Chen, Chin. Med., 2020, 134, 783e791.

3. L. Zhang, F. Ren, Q. Zhang, Y. Chen, B. Wang and J. Jiang, Dev. Cell, 2008, 14, 377–387.

4. Y. Zhao and X. Yang, Int. J. Cancer, 2015, 137, 2767–2773.

5. J. V. Patil, S. S. Shubhangi, U. Shweta, G. Pankaj and B. Suresh, ChemistrySelect, 2024, 9, e202303391.

6. S. M. Emam, B. Samir and A. M. Ahmed, Results Chem., 2023, 5, 100725.

7. A. N. Ahmed, H. A. Aboubakr and M. S. Maha, Bioorg. Chem., 2023, 131, 106299.

8. A. N. Ahmed, A. S. Ihsan, H. A. Aboubakr and M. S. Maha ACS Omega, 2022, 7, 25265– 25277.

9. M. Islam, M. Abdullah, N. Essam, Y. Sammer, A. Muhammad, N. Asif and W. Abdul, J. Mol. Struct., 2022, 1269, 133843.

10. S. Dabhade, P. Manjushri, S. Lala, A. Sachin, A. Shweta, Y. Somdatta and N. Santosh, Chem. Biodiversity, 2024, 21, e202400015.

11. F. Rizk, N. Ahmed, A. Seham and K. Amira, J. Heterocycl. Chem., 2022, 59, 2190–2206.

12. H. S¸enol, G. Mansour, ¨O. Gülbahar, T. AlimandGU˘gurJMol.Struct.,2024,1295,136804.

13. A. N. Ahmed, A. A. El-Barbary, M. S. Maha, A. Hayam and S. Mohamed, Sci. Rep., 2024, 14, 3530.

14. K. J. Setshedi, M. B. Richard, I. Kayhan, M. Dorien, C. Guy and J. L. Lesetja, Med. Chem. Res., 2024, 3, 977–988.

15. A. N. Ahmed, A. I. Saham, A. Hayam, M. A. Nora andS. A. Hamada, J. Iran. Chem. Soc., 2023, 20, 2963–2976.

16. S. S. Sha, Indian J. Chem., Sect. B: Org. Chem. Incl. Med. Chem., 2021, 60, 1132-1136.

17. S. K. Ramadan and A. R. Sameh, J. Iran. Chem. Soc., 2022, 19, 187–201.

18. J. Zhou, W. Feng, X. Bin, L. Xin, P. Cheng and F. U. Peng, Oncol. Res., 2024, 32, 943.

19. H. Tanaka, N. Seito, O. Kenji, T. Tomoya and H. Toshihiko, Biochem. Biophys. Res. Commun., 2009, 381, 566–571.

20. G. Padmavathi, K. R. Nand, B. Devivasha, A. Frank, M. Srishti, S. Gautam, B. Anupam and B. K. Ajaikumar,Phytomedicine, 2017, 25, 118–127.

21. A. Jasiecka, T. Maslanka and J. J. Jaroszewski, Pol. J. Vet. Sci., 2014, 17, 207–214.

22. H. G. Kraetsch, T. Hummel, J. Lötsch, R. Kussat and G. Kobal, Eur. J. Clin. Pharmacol., 1996, 49, 377–382.

23. A. Weissberger, H. D. Porter and W. A. Gregory, J. Am. Chem. Soc., 1944, 66, 1851–1855.

24. J. Dong, H. Guang, Z. Qijing, W. Zengtao, C. Jiahua, W. Yan, M. Qingqing and L. Shaoshun, MedChemComm, 2019, 10,1606–1614.

HISTORICAL AND ARCHEOLOGICAL SCIENCES

ON THE ISSUE OF THE ECONOMIC ACTIVITY OF THE SHIRVANSHAH DYNASTY IN THE HISTORICAL ASPECT DURING THE EARLY MIDDLE AGES

Aliyev F.
PhD in Historical Sciences, Associate Professor, Azerbaijan State Oil and Industry University, Department of Human Sciences
Baku, Azerbaijan
DOI: 10.5281/zenodo.21617827

Abstract
This work addresses a topical research. The expanding scope of the research is evident throughout the presentation. Although the article clearly has a strong historical focus, we believe that the author's chosen perspective will undoubtedly be of interest to representatives of related professions. This is urged by elementary logic: the history of the Shirvanshah Empire is inseparable from the conquests of foreign invaders. Consequently, it is closely linked to the politics of that country; at the same time, economic activity affects certain cultural strata of the population. Consequently, it may be considered relevant for study by historical scientists, cultural scholars, philosophers, politicians, political scientists, and linguists. All of the above (as is readily apparent) is contained within this work. The article is clearly divided into three parts, which clearly demonstrates the author's in-depth and thorough examination of both the critical and specialized (i.e., historical) issues under investigation. Specifically, the article begins with a brief but essential historical background on the emergence of the Shirvanshah state and the ancient and medieval scholars and philosophers who wrote about it and the Azerbaijani cities adjacent to it. Next, the central theme is directly addressed. This is followed by information directly on the development and improvement of all parameters of the economic system in this country. Significantly, this activity is presented inextricably linked to the politics, ideology, and culture of Shirvan. This multilayered connection runs throughout the entire work. This structure clearly demonstrates the author's serious approach to the subject under analysis. It is also gratifying that not a single historical event or theoretical proposition is left without definitive conclusions or additional commentary.
Keywords: Shirvanshah dynasty, Medieval Azerbaijan, economy, politics, handicraft production, trade relations, international relations.

Introduction. Research into the various aspects of the existence and development of the Shirvanshah Empire has long attracted the attention of scholars representing various disciplines. Suffice it to say that it has been addressed from various angles by more than just historians. First and foremost, we hasten to point out that the very history of this state is as marked by its military and political environment as it is by its dramatic history. For this reason, most researchers in the 20th century addressed its military doctrine, economics, politics, and related economic activities.

The wide range of scholarly approaches to this topic is entirely justified by the circumstances. It is significant that chroniclers and ordinary travelers naturally and naturally joined in the scholarly pursuits over many decades. Romantics were typically drawn to the exoticism of the Caucasus; more serious critics, to the heroism of this typically feudal state. Meanwhile, people of diverse views (as evidenced by numerous works) were drawn to its history. It's significant that the ancient philosopher Ptolemy, in the second century AD, wrote about Shemakha (under the then name "Shemakhia"), referring to this city, as modern experts argue, as "precisely Shirvan with its future capital" [2, p.378].

Literature review. A millennium later, we find information about this state dating back to the Early Middle Ages. Thus, the politics, economics, ideology, a series of military events, and agriculture of Shirvan are described in Arabic-Iranian manuscripts. In our time, renowned Russian and Soviet historians and philosophers I. Krachkovsky, V. Minorsky, and I. Petrushevsky have taken advantage of this information. Of particular interest are the works of Dagestani scholars R. Shikhsaidov, V. Kotovich, and several others. The scholarly works of Azerbaijani researchers such as A. Alizade, A. Bakikhanov, Z. Bunyadov, O. Ismizade, I. Jafarzade, S. Ashurbeyli, and others are indicative. Consequently, the keen interest in various aspects of the development of the Shirvanshah state today is entirely justified.

Results and discussion. A historical review of the Early Middle Ages (the turn of the 10th and 11th centuries) allows us to establish that the Shirvanshah dynasty wielded real political power after the collapse of the Abbasid Caliphate. Moreover, during this same historical period, this dynasty also took control of the economy. The political situation, in turn, stabilized, as the Shirvanshahs gained autonomy and independent governance. Significantly, a genetic factor was added: power, passing from hand to hand, was hereditary. Thus, economic activity became subordinate to ideology and politics. Clarifying how this chain of events unfolded in the Shirvanshah state during this period of their reign from a historical perspective is the main goal of this article.

First of all, we emphasize that any economic activity cannot develop at full capacity in a territory periodically subject to invasions. This factor had to be taken into account. The point was that Baku, Shirvan, and Shamakhi were very wealthy cities during the period we've just mentioned. This naturally attracted the attention of foreigners. Modern historians point out that

invaders repeatedly attacked the Shirvanshah state, plundering, destroying agricultural lands, and taking its inhabitants captive. For example, bare statistics tell us of one of the bloodiest Rusich invasions in 1029-1030 near the territory of present-day Baku, which claimed a large number of Shirvan victims.

Thirty years later (more precisely, by 1065-1066), the eastern tribes, led by Karatekin, also attacked Shirvan, successively conquering Shamakhi and then Baku. In the second half of the 11th century, not only the Shirvanshah state but the vast territory of Transcaucasia in general was subjected to periodic attacks by several Turkic-speaking tribes. During this period, the Seljuk army was particularly powerful. Its dynasty consisted not only of Turks but also of other nationalities. Since ancient times, the Seljuks had always been considered brave and powerful warriors. The ever-increasing pressure from foreign conquerors of various nationalities naturally and logically disrupted not only the established, measured rhythm of life in the Shirvan state but also, to a certain extent, dismantled the entire economic complex. The problem was further exacerbated by the fact that in the second half and very end of the 11th century, most of the Shirvanshahs' kingdom led a sedentary lifestyle. Therefore, consolidating the economy into a single unit was extremely difficult. It was essential to simply concentrate the local population within a specific area of Shirvan. This dispersal, in turn, led to a general decline in the economy. Thus, throughout the 11th century, the economy of the Shirvanshahs suffered significant damage, because attacks from various armies continued. However, in fairness, it should be emphasized that even in such a difficult situation, the Shirvan authorities strove, as far as possible, to preserve their independence and democratization.

Clearly, such frequent raids on the Shirvanshahs' realm not only affected agriculture, but, above all, disrupted the peaceful flow of life. There were instances in the history of this country where isolated, destructive raids by the enemies of Shirvan lands would turn into protracted, drawn-out battles. These could cause devastating damage to its natural environment. To the credit of the Shirvan people, it should be noted that at the turn of the 11th and 12th centuries, thanks to a sound state strategy, coupled with the hard but productive labor of ordinary citizens, Shirvan gradually recovered, and its economic activity generally strengthened.

More detailed information about this period can be found both on Internet portals and in relevant historical chronicles of that turbulent time for the Shirvanshah state. Given the thoroughness of this study, we deem it unnecessary to present historical facts in strict sequence throughout the 11th century. More important for us is the reference to the 12th century, when the Eldegizid dynasty, comprising the population of Southern Azerbaijan, continued its relentless raids on Shirvan, capturing the capital. Only in the last quarter of the 12th century did the ruler of the Shirvanshah dynasty, Akhsitan I, with the active support and assistance of George III, manage to deal with the Rusichs fleet with dignity. From that time on, partly under the flag of the Eldegizids, the Shirvanshah state nevertheless gained long-awaited independence, but the forced democratization of the country could not help but negatively impact its economic activity as a whole. Thus, it is clear that the development and improvement of economic activity and its overall way of life in the state of the Shirvanshahs were primarily negatively impacted by periodic (or, to be more precise, incessant) raids by foreign enemies, which resulted in the destruction of its agriculture.

However, there is also a positive side. This state has always been favorably located geographically. Specifically, its territory stretched from the "Wild Kura" all the way to Derbent. Mountainous areas were relatively few here. Therefore, the lands located in this area were historically very fertile. For centuries, the Shirvan people intensively engaged in livestock breeding and agriculture. Moreover, various trades and crafts flourished everywhere. In this regard, the famous Azerbaijani historian Sara Ashurbeyli, who for many years specially studied various political, ideological, cultural and commercial aspects of the Shirvanshahs' state, records in one of her works that in the "15th-16th centuries, handicraft production was especially widespread in the capital of Azerbaijan and on the Absheron Peninsula" [1, p.89]. According to her, there were many professional metal and stone carvers in Shirvan, the state was also famous for its engravers, poets, artists, calligraphers, and weavers. The art of carpet weaving flourished, jewelers made amazing jewelry from gold and other precious metals (as if to the envy of their enemies), architects actively worked on urban development, and potters worked diligently. Shipbuilding occupied an important place among handicraft industries, etc. In addition to the above, historians point out that agricultural lands also served as a reliable foundation for the development of domestic production. Consequently, alongside male agricultural labor, female labor was also used to some extent. The negative impact of military attacks was, so to speak, counterbalanced by the high level of agricultural labor.

This state of affairs facilitated very early settlement. Even in the early medieval period, settlers were able to successfully combine agricultural and industrial activities. This development strengthened trade relations and cultural ties with other nations and tribes. Gradually, the Shirvanshah state freed itself from the dominance of the former tribal communities and organically evolved new feudal relations. Certainly, this was a significant step forward in political, ideological, and economic terms. Objectively speaking, progress consisted in the fact that the new formation in Shirvan seemed to provoke the accumulation of residual capital. Naturally, this capital was concentrated in the hands of the wealthiest segment of the local population. But these individuals, figuratively speaking, set the tone, and ultimately, the wealth of individual Shirvan patrons smoothly and naturally led to the emergence of Shirvan's richest cities. In any case, most of our modern researchers confirm this fact through their comparative analysis of the Shirvanshahs' state with other Eastern and even some Western European countries of the period under consideration. To this, of course, are added the natural resources we've already described, coupled with the rise of productive forces in Shirvan and the development of trade relations.

Archaeological excavations in recent years have shown that favorable geographic and ethnographic factors have also made it possible to discover such natural resources as oil production, salt lakes, and valuable non-ferrous metals and minerals in Shirvan. Moreover, the Shirvanshahs' carefully thought-out strategy led to rapprochement with several neighboring countries, with which good-neighborly relations were established. This is particularly confirmed by the historian, political scientist, and cultural scholar A.O. Murtazayev, who speaks of the stable economic and sociopolitical position of this state, simultaneously cemented by cultural ties with Dagestan [4, p.5-15].

We have already written above that the Shirvanshahs' state partially attracted the attention of people of diverse views and beliefs. For example, its Caucasian exoticism attracted both foreign conquerors and ordinary, inquisitive people interested in the country's historical roots - quite peaceful people who migrated from adjacent eastern regions to Shirvan to engage in productive agricultural work. Thus, over a long period of time, this state gradually became more densely populated, and its ethnic group accordingly became more diverse.

V.F. Minorsky [3, p.110-113], in particular, points to this fact. Moreover, this prolific scholar, as his primary evidence, wisely undertakes an extensive historical review of the issue under study. This history is quite extensive in V.F. Minorsky's monograph, so we will present individual excerpts in the most generalized form necessary for the confines of this article.

From the words of this historian, it follows that the culture of the Shirvanshah state was shaped by its origins in ancient times. In the period before the Early Middle Ages in the East, the population of this region consisted primarily of aborigines. However, even in those distant years, the local population was distinguished by its wealth compared to many other eastern regions. And it turns out, this wealth was due to the observance of their traditions and agricultural way of life. In any case, this wealth did not come at the expense of the country's political strategy. It should be noted that this activity subsequently consisted of relatively primitive handicrafts and industrial production. But soon, the residents of Shirvan and the surrounding towns began to adopt not only the economic practices themselves but also the cultural demands of other regions. The close mutual influence of several cultures allowed the agriculture of the Shirvan state to elevate its own to an incomparably higher level. Moreover, it was filled with such technological innovation that historians began to record, without exaggeration, a qualitative leap.

There is another noteworthy point that cannot be overlooked. The state of the Shirvanshahs in the Late Middle Ages was distinguished by its national diversity, consisting of a number of ethnic groups. Each of them was autonomous, relatively independent of the others, and therefore had a real opportunity to make a modest contribution to the overall cultural treasury of the Shirvanshahs. And it was extremely rich. In other words, the Shirvan Khanate during the historical period we have described was divided into several regions. Each of them was a kind of "mini-state" or "state within a state". And this is not an empty statement. There are reliable examples of this. Thus, the historian T.M. Shavlayeva points out in one of her articles that "the greatest masters of the arts of the Shirvan state, over the course of several centuries, with their skillful hands created genuine masterpieces of architectural ensembles, literary monuments, jewelry, and carpets" [6, p.148-149]. This corresponds to reality. With regard to the latter art form, it is all the more gratifying to emphasize that the state of the Shirvanshahs for many decades firmly gained popularity throughout the world. Carpet weaving occupies a special place here, having received high marks from experts and prestigious awards at various international exhibitions in recent years. For example, carpets from Shamakhi and neighboring countries enjoyed worldwide fame. And so on.

With such vibrant economic performance, the 12th century of the Shirvan state bore long-awaited and fruitful fruit. Specifically, up until 1100, there was a significant upsurge in this area. Moreover, agricultural work was carried out here with the same vigor and persistence as handicraft production. This soon impacted the trade relations of the Shirvanshah states with other countries in Asia and Europe. The very boundaries of Eastern cultural and trade relations expanded. This state began to import more of its goods abroad: the production of silk, wool, linen, and carpets became increasingly important. While earlier in this article we wrote about the episodic role of craftsmen, from now on (from 1100 onward), the production of metalwork, pottery, and faience products increased significantly. In other words, economic activity in various sectors, in modern parlance, was put on stream. Moreover, the expansion of trade and economic ties increased the power of the Shirvan fleet. This not only strengthened state borders but also had a beneficial effect on economic affairs. Shirvan, for example, began exporting increasingly fresh and dried fruits, fish, oriental sweets, and salt, and oil production was established. The latter, although in a very primitive form by today's standards, also contributed to the overall economic upswing.

An interesting reference is provided by a publication called "Soviet Union". From the 1960s to 1980s, it was published as a series of essays on all the republics of the former USSR. (The publication ceased publication only in the early 1990s due to the collapse of the Soviet Empire.) Under the corresponding section—"Azerbaijan" - for 1971, an operational report on the production of raw silk, rice products, the number of sheep and horses in the state of the Shirvanshahs was also published. Agricultural products such as "mulberry trees, pomegranates, grapes, and cotton" are noted [5, p.67]. Significantly, the summary we cited concluded with the editors' assurance that, in the state of the Shirvanshahs, along with a general and significant economic upswing, the socio-political life of the people also changed for the better.

It is regrettable to note that this glorious state subsequently suffered numerous misfortunes and social catastrophes. There is nothing surprising or paradoxical about this. As the well-known proverb goes, a tasty morsel, like forbidden fruit, is always sweet. In the first half of the 13th century, Shirvan was successively attacked first by the Tatar-Mongol armies, then by Timur's warriors. The previously flourishing cities of the

Shirvan zone were literally reduced to ruins. Only by the mid-15th century did the state gradually begin to strengthen its agricultural, trade, and economic base. But this is a different era, a different social formation, which is beyond the scope of this article.

Conclusions. This article focuses on a single, overarching aspect: economic activity in the Shirvanshah state. It is primarily presented as a dynamic process that developed and gradually improved in this region, alongside the aforementioned political events and social ideology. Thus, over the course of several centuries, this activity is demonstrated as the result of carefully considered tactics and strategies by the state's ruling structures. For clarity, economic activity in the state is divided according to various specialties. This means that every resident of this country, engaged in a specific production, contributed to the overall development of this region of the planet. This type of activity is described as an important cultural component of the Shirvanshah state. Its development was actively facilitated, on the one hand, by the skilled hands of laborers, which is recognized as a merit and distinctive trait of the Azerbaijani people. On the other hand, by a favorable geographical location.

References

1. Ашурбейли С.А. История города Баку. Баку: Азернешр, 1992, с. 87-179.
2. Большая Советская Энциклопелия, 1978, т. 29, 469 с.
3. Минорский В.Ф. История Ширвана и Дербенда. М., 1963. 266 с.
4. Муртазаев А.О. Политическое положение Кайтага в XVI в. // Вестник Института ИАЭ ДНЦ РАН. 2014. № 1. С. 5-15.
5. Советский Союз. Азербайджан. М., 1971. 318 с.
6. Шавлаева Т.М. Хозяйственная деятельность в государстве Ширваншахов. // Культура и цивилизация, 2018, т. 8, № 2. с. 145-150.

ЖЕНСКИЕ ЛИЦА ИЗ НАХИЧЕВАНИ СРЕДНИХ ВЕКОВ

Алиев А.А.
Руководитель проекта ДНК Азербайджана, лаборатория FTDNA.
ORCID ID: 0000-0002-3729-815X
Керимов З.М.
Кандидат медицинских наук, доцент департамента биологических наук, Университет «Хазар»
ORCID ID: 0000-0002-9438-3721
Евсеев Р.В.
Старший научный сотрудник Лаборатории реконструкции внешности «Other worlds studio».
ORCID ID: 0000-0001-8826-3961
Ибрагимов А.Ш.
Кандидат медицинских наук, доцент департамента биологических наук, Университет «Хазар»
ORCID ID: 0000-0001-6426-3246

FEMALE FACES FROM MEDIEVAL NAKHCHIVAN

Aliyev A.
Azerbaijan DNA Project Manager, FTDNA Laboratory.
ORCID ID: 0000-0002-3729-815X
Kerimov Z.
PhD in Medicine, Associate Professor of the Department of Biological Sciences, Khazar University
ORCID ID: 0000-0002-9438-3721
Yevseyev R.
Senior Researcher, Laboratory of Appearance Reconstruction, "Other Worlds Studio."
ORCID ID: 0000-0001-8826-3961
Ibragimov A.
PhD in Medicine, Associate Professor of the Department of Biological Sciences, Khazar University
ORCID ID: 0000-0001-6426-3246
DOI: 10.5281/zenodo.21617841

Аннотация

Были исследованы женские черепа, обнаруженные археологами при раскопках средневекового городища Хараба-Гилан (Ордубадский район, Азербайджанская Республика). Краниометрические и краниоскопические данные позволили определить пол, расу, возраст и некоторые индивидуальные особенности исследованных черепов. Эти данные использовались для внесения корректировок в итоговые изображения, которые были получены при реконструкции лиц людей, кому эти черепа принадлежали. Визуализация внешнего облика женщин (чьи черепа исследовались) осуществлялась в Лаборатории реконструкции внешности «Other worlds studio» на основе модернизированного метода М.М. Герасимова. Итоги работы наглядно демонстрируют преемственность современных жителей с популяциями, жившими на территории средневекового Азербайджана. Результаты работы имеют междисциплинарный характер и представляют интерес как для археологов и историков, так и для специалистов занимающихся физической антропологией, морфологией человека и этнографией.

Abstract

Female skulls discovered by archaeologists during excavations at the medieval archaeological site of Kharaba-Gilan (Ordubad district, Azerbaijan) were analyzed. Craniometric and cranioscopic data allowed us to determine the sex, race, age, and some individual characteristics of the skulls. These data were used to adjust the images obtained for facial reconstruction of the women whose skulls were studied. The appearance of these women was visualized in the «Other worlds studio» Facial Reconstruction Laboratory using a modernized method developed by M.M. Gerasimov. The results clearly demonstrate the continuity between modern residents and the population that inhabited the territory of medieval Azerbaijan. The results are interdisciplinary in nature and are of interest to archaeologists and historians, as well as specialists in physical anthropology, human morphology, and ethnography.

Ключевые слова: краниология, реконструкция лица по черепу, средневековый Азербайджан, историческая антропология.

Keywords: craniology, facial reconstruction from skulls, medieval Azerbaijan, historical anthropology.

Введение

В последние годы у азербайджанских исследователей возрос интерес к реконструкции внешности древних жителей различных регионов Республики [1, 2]. Во многом этому способствовало появление современных компьютерных технологий и программ (в том числе систем компьютерного зрения, на основе искусственного интеллекта), значительно облегчающих графическую визуализацию внешности человека по его черепу. Однако, количество завершённых реконструкций пока незначительно,

что по большей части обусловлено отсутствием должного внимания к этой тематике со стороны археологов. Ввиду этого нами были проведены исследования с целью воссоздания прижизненного облика нескольких женщин, живших на территории современной Нахичевани в эпоху государства ильханидов (13-14 века), которые были обнаружены при раскопках городища Хараба-Гилан.

Материалы и методы

В исследовании использовались женские черепа, найденные учёными из группы Хараба-Гиланской археологической экспедиции Института археологии и антропологии НАНА. Всего при раскопках погребений в городище Хараба-Гилан (Ордубадский район, Азербайджанская Республика) было обнаружено 14 черепов мужчин и женщин. Археологи, исходя из анализа артефактов, характера погребения и летописных сведений отнесли примерное время жизни этих людей к периоду 13-14 веков, когда территория современной Нахичевани входила в состав государства Хулагидов, пятого улуса распавшейся монгольской империи Чингисхана [3].

Первичные краниометрические и антропологические исследования были проведены сотрудником НАНА Кириченко Д.А. [4]. Несмотря на неполный охват краниометрических параметров и краниоскопических признаков, которые бы дали возможность провести диагностику антропологического типа индивидов из погребения, прогнозировать их соматотип, а также получить медико-криминалистичекое описание внешности, имевшееся количество фото- и видеоматериалов экспедиции позволило осуществить визуализацию внешности по черепу в изометрической проекции.

Реконструкция внешнего облика людей по их черепам осуществлялась в Лаборатории реконструкции внешности «Other worlds studio» на основе модернизированного метода М.М. Герасимова [5].

Результаты исследования и обсуждения

Как мы отметили, при работе с указанными черепами мы не располагали всеми необходимыми краниометрическими и краниологическими данными для краниофациальной реконструкции. Соответствующие черепа пока законсервированы на территории городища Хараба Гилан, где создаётся археологический мавзолейный комплекс. Тем самым, мы пока не смогли измерить черепа, проставить соматометрические точки, а также провести цефалорадиографию и компьютерную томографию черепов. Для каждого отдельного кейса не удалось диагностировать рост субъектов, конституцию (соматотип телосложения индивидов), антропологический тип. Эти данные позволяют существенно корректировать конечное изображение, полученное методом фотореконструкции. Как правило, эти данные и результаты ДНК-анализов костных останков в совокупности дают представление о признаках фенотипа (цвет глаз, волос и кожи, степень развития мускулатуры, пропорции тела и пр.). Тем не менее, определённые корректирующие сведения предварительно у нас всё таки имелись. Так, археологами были диагностированы: половая принадлежность черепов, примерный возраст, описаны некоторые краниоскопические признаки, выявлены деформации, а также следы травм и болезней [4].

Имеющиеся в распоряжении данные давали возможность высказаться, что все исследованные черепа принадлежали женщинам. Краниоскопические признаки (облитерация швов) и оценка стёртости зубов давала возможность высказаться о приблизительном биологическом возрасте исследуемых черепов. Биологический возраст первой женщины на момент смерти оценивался примерно в 25-35 лет (северный сектор мавзолея, череп №9). Возраст второй женщины соответствовал приблизительно 17-20 годам (северный сектор мавзолея, череп №10). И, наконец, третьей женщине на момент смерти было примерно 40-45 лет (южный сектор мавзолея, череп №8). На всех черепах сохранились некоторые зубы, что предопределяет некоторый оптимизм относительно результативности генетических исследований в будущем. Череп №8 из южного сектора мавзолея был полным краниумом, тогда как на двух других черепах нижняя челюсть отсутствовала. Хотя на каждом черепе были измерены только несколько краниометрических признаков, эти данные позволяли достаточно точно отнести этих женщин к представителям южной ветви большой европеоидной расы [6].

Некоторые краниоскопические признаки (форма мозгового отдела черепной коробки, ширина и профилировка лицевого отдела, выступание носа, формы орбит и др.) зафиксированные антропологпми во время полевых работ [4], также, позволяли высказаться о некоторых индивидиуальных признаках тех людей, чьи останки стали объектом изучения учёных из археологической экспедиции. Кроме того, на одном из черепов отмечается кистоидный дефект нижней челюсти, что скорее всего сопровождалось эстетическим искажением внешнего облика. Эта особенность учитывалась при внесении корректировок в изображение, полученное при реконструкции внешности по черепу в изометрической проекции. Однако, другие патологические признаки, отмеченные во время первичного исследования черепов, не являются индивидуализирующими и характерны для популяции в целом [7]. Остеомы на костях свода черепа находятся в зоне покрытия капиллиция и могут не учитываться при реконструкции внешности.

Таким образом, в результате анализа краниометрических и краниоскопических признаков трёх женских черепов из городища Хараба-Гилан были получены определённые данные по морфологическим признакам внешности этих индивидов (чьи черепа исследовались). Эти сведения использовались для внесения соответствующих поправок к фотоизображениям, полученным при визуализации внешнего облика этих в Лаборатории реконструкции внешности.

Пигментация кожи, цвет глаз и волос условные и выбраны на основе расовой принадлежности исследуемых черепов и поэтому результаты работы по восстановлению лица трёх жительниц городища Хараба-Гилан на данном этапе следует считать

предварительными. Более точные данные по пигментации кожи, цвету волос и глаз, а также по некоторым анатомо-морфологическим особенностям элементов физиономии можно будет получить после секвенирования ДНК из этих костных останков и проведения более объёмной краниометрии (она позволит осуществить подробное медико-криминалистическое исследование черепов).

Фотографии черепов и соответствующие реконструкции внешнего облика тех женщин, кому они принадлежали представлены на рисунках 1-3 ниже.

Рис. 1. Череп и реконструкция внешности женщины 25-35 лет

Рис. 2. Череп и реконструкция внешности женщины 17-20 лет

Рис. 3. Череп и реконструкция внешности женщины 40-45 лет

Судя по выполненным реконструкциям можно предположить, что все три женщины внешне немного похожи. В данном случае, учитывая место обнаружения их останков и особенности захоронения, можно предположить, что эти индивиды, возможно, были родственницами. Однако, как мы отметили, для научно-обоснованного ответа на этот вопрос необходимы ДНК-исследования.

Вместе с тем, в облике этих женщин прослеживается преемственность с современным населением Азербайджана, наблюдается близость к обобщённым местным типажам. В рамках различных расовых классификаций физиономические особенности наиболее схожи с описанием представителей индоевразийского локального расового типа, выделяемого М.Г. Абдушелишвили [8]. Аналогичное сходство прослеживается и с каспийским антропологическим типом, который выделял В.П. Алексеев [9]. Здесь уместно сказать, что каспийский антропологический тип считается самым распространённым среди современного населения Азербайджана и это отмечалось как в антропологических [10], так и краниологических исследованиях [11].

Выводы

Результаты работы по реконструкции лица по черепу трёх женщин, чьи костные останки стали объектами изучения антропологов из Хараба-Гиланской археологической экспедиции, являются продолжением исследований по визуализации внешнего облика людей из исторического прошлого Азербайджана и очень важны в контексте популяризации исторической антропологии. Впервые получена графическая визуализация внешнего облика жителей средневековой Нахичевани, и эти результаты наглядно демонстрируют преемственность современных жителей с популяциями древности, жившими на территории Азербайджана. В целом итоги работы должны усилить интерес к соответствующей тематике не только специалистов, занимающихся археологией и антропологией непосредственно (в связи с профессиональной деятельностью), но и учёных из смежных научных отраслей.

Благодарности

Авторы выражают благодарность руководителю Хараба-Гиланской археологической экспедиции Института археологии и антропологии НАНА, доценту Б.И. Ибрагимли за предоставленные фото- и видеоматериалы, а также за ценные устные консультации.

References

1. Ibrahimov A.S., Yevseyev R.V., Aliev A.A. Forensic facial reconstruction of a man from the Shamakhi gravefield of the early medieval period // Moscow University Anthropology Bulletin (Vestnik Moskovskogo Universiteta. Seria XXIII. Antropologia), 2023; 1/2023; c. 129-137 (in Russian). DOI: 10.32521/2074-8132.2023.1.129-137

2. Huseynov V.K., Musayev Sh.M., Ibragimov A.Sh., Yevseyev R.V., Kerimov Z.M., Aliev A.A. Reconstruction of the Appearance of a Woman who lived on the Absheron Peninsula around the 14th-17th centuries. Ann Rev Resear. 2025; 13(2): 555857. DOI: 10.19080/ARR.2025.13.555857

3. Nurieva I.T. History of Azerbaijan: from ancient times to the present day (in Azerbaijani). Baku, “Science and Education” Publishing House, 2022, 744 p.

4. Кириченко Д.В. Городище Хараба Гилан. Мемориальный комплекс II. Мавзолей I. Северный и южный секторы. Отчёт Отдела «Антропология» ИАА НАНА, 2025 г., с. 120-165.

5. Veselovskaya E.V. Anthropological reconstruction of human appearance. Development and application of new methodological approaches. Dissertation for the degree of Doctor of Historical Sciences, Moscow, 2016, 194 p. (in Russian).

6. Elliott M.N., Fremont A., Morrison P.A., Pantoja P., Lurie N. A new method for estimating race/ethnicity and associated disparities where administrative

records lack self-reported race/ethnicity. Health Services Research, 2008, Vol43(5), p. 1722-1736. DOI: 10.1111/j.1475-6773.2008.00854.x

7. Ortner D.J. Identification of pathological conditions in human skeletal remains. London: Academic Press, 2003, 645 p.

8. The origins of the formation of the modern population of South Asia. Moscow, “Nauka” Publishing house, 1990, 411p. (In Russian).

9. Alekseev V.P. Selected works in 5 volumes. Nauka, Moscow, 2009, Vol5, p. 91-92. (In Russian).

10. Kasimova R.M. Anthropological studies of the modern population of the Azerbaijan SSR, Baku, “Elm” Publishing house, 1975, p. 100 (In Russian).

11. Musayev Sh.M. Forensic medical craniological identification of the inhabitants of the Caucasus. Baku, “Elm” Publishing house, 1999, p. 130.

MEDICAL SCIENCES

THE POTENTIAL ROLE OF SERUM VITAMIN D LEVEL IN MIGRAINE HEADACHE: A CASE–CONTROL STUDY

Shkiryak A.
CEO and chief medical officer "Shkiryak Clinic"
Ukraine
Neurosurgeon
MD, MsSc
ORCID: https://orcid.org/0000-0002-2282-8285
Khanenko S.
"Shkiryak Clinic"
Ukraine
Family doctor
MD, MBA
ORCID: https://orcid.org/0009-0000-4237-1158
DOI: 10.5281/zenodo.21617862

Abstract

Purpose: Much concern was directed toward exploring the relationship between vitamin D and migraine. There is strong evidence that vitamin D supplementation can decrease frequency, severity, and duration of migraine headache attacks. The aim of this work was to investigate the difference in serum levels of 25 (OH)-vitamin D between patients with migraine and healthy controls, to determine the differences in headache characteristics according to vitamin D status, and to correlate serum 25 (OH)-vitamin D level with duration, frequency, and severity of migraine headache attacks.

Patients and methods: This is a case–control study conducted on 40 patients diagnosed with migraine and 40 healthy controls. History was taken from patients with migraine regarding headache characteristics. Migraine severity scale (MIGSEV) and Headache Impact Test-6 (HIT-6) were used for migraine assessment. Serum 25(OH)-vitamin D was measured for all patients and controls using enzyme-linked immunosorbent assay (ELISA).

Results: Patients with migraine had significantly lower 25(OH)-vitamin D serum level in comparison to controls (*P*-value=0.019). The incidence of aura, phonophobia/photophobia, autonomic manifestations, allodynia, and resistance to medications were significantly higher in migraineurs with vitamin D deficiency than those with normal vitamin D. There was a statistically significant negative correlation between 25(OH)-vitamin D serum level and attack duration in hours (*P*-value<0.001), frequency of the attacks/month (*P*-value<0.001), MIGSEV scale (*P*-value=0.001), and HIT-6 scale (*P*-value=0.001).

Conclusion: Patients with migraine had significant vitamin D deficiency compared to healthy controls. Such deficiency significantly affects headache characteristics, duration, frequency, and severity of headache attacks.

Keywords: MIGSEV, HIT-6, 25(OH)-vitamin D, ELISA.

Introduction

Migraine and tension headache are the most common primary headache disorders that affect 80% of the people all over the world.[1] Global estimation of migraine headache prevalence showed that migraine affects 1 in 10 people worldwide.[2] It affects 6% of the men and18% of the women and has a peak incidence in ages between 25 and 55 years.[3] A systematic analysis for the Global Burden of Disease (GBD) Study in 2015 considered migraine headache as the second largest contributor of disability-adjusted life years (DALYs) among all neurological disorders.[4]

Migraine attacks are characterized by being moderate-to-severe, pulsating, unilateral, associated with nausea, vomiting, phonophobia, and photophobia. The attack usually lasts from several hours to 2–3 days.[5] About 25% of the patients with migraine perceive an aura, which is a transient disturbance in visual, sensory, language, or motor function preceding the migraine attack. Migraine headache is caused by the release of pain-producing inflammatory mediators around the cranial nerves and blood vessels. Such mediators induce vascular smooth muscle dysfunction.[6,7]

Options for migraine treatment remain unsatisfactory because of the reported lack of effectiveness and the significant side effects. Thus, efforts to identify more effective well-tolerated therapy for preventing migraine remain urgent.[8]

In recent years, vitamin D deficiency has been reported as a global public health problem. Prevalence of vitamin D deficiency ranged between 30% and 50% in normal populations.[9,10] Despite high sun exposure in middle eastern countries, these countries are considered among the highly prevalent areas for vitamin D deficiency in the world.[11]

There is strong evidence supporting an association between vitamin D deficiency and chronic pain.[12] Much concern was directed toward the presence of a possible relationship between vitamin D and migraine.[13–15] Vitamin D was found to have a role in the pathways involved in the pathogenesis of migraine including pain sensitization, inflammation, and immune dysfunction.[16,17] On the molecular level, Motaghi et al revealed that vitamin D receptor (VDR) gene polymorphisms may increase the risk for developing migraine without aura.[18]

Aim of this work

The aim of this work was to investigate the difference in the serum levels of 25-hydroxy vitamin D (25(OH) vit D) between patients with migraine headache and healthy controls. The secondary objectives were to determine the differences in headache characteristics according to vitamin D status and to correlate serum 25(OH)-vitamin D level with duration, frequency, and severity of migraine headache attacks.

Materials and methods

Study design

The present study is a case–control study. It was conducted on 40 patients diagnosed as having migraine headache with or without aura and 40 normal healthy controls. The patients were recruited in the period between October 2017 and October 2018 from the Neurology outpatient clinic and Pain clinic in Beni-Suef University Hospital, Egypt. A written informed consent was obtained from each participant in this study or from parent or legal guardian for participants under the age of 18 years. The study was conducted in accordance with the Declaration of Helsinki. The study was approved by local ethical committee in Faculty of Medicine, Beni-Suef University, Egypt. The committee's reference number is FWA00015574.

Inclusion criteria

Selected patients fulfilled the criteria for diagnosis of migraine headache based on International Classification of Headache Disorders-II (ICHD-II) diagnostic criteria.[7] The age of the selected patients ranged between 15 and 55 years.

Exclusion criteria

Patients in the migraine group were excluded from the study if they have secondary headache, bilateral papilledema, or magnetic resonance imaging (MRI) showing structural brain lesion. Patients with a history of concomitant medical or metabolic illness known to affect vitamin D level, such as infectious disorders, liver or kidney disease, gastrointestinal disease, cancer, sarcoidosis, or tuberculosis, were also excluded. Participants in both groups were excluded from the study if they consumed vitamin D supplements in the preceding 3 months (any dose); or if they were taking medications that could affect vitamin D serum level such as glucocorticoids, thiazide diuretics, or statins. Pregnant were also excluded from the study.

Participants of this study were subjected to the following

History taking regarding

The frequency of headache attacks/month, the duration headache attacks, the presence of aura, nausea/vomiting, phono/photophobia, autonomic manifestations, allodynia, and resistance to medications.

Migraine assessment

It consists of six items: pain, role functioning, social functioning, vitality, psychological distress, and cognitive functioning. The patient has to answer each of the six questions using one of the following five responses: "always", "very often", "sometimes", "rarely", or "never". The total HIT-6 score ranges between 36 and 78, where a higher score of the test indicates a greater impact of migraine headache on daily activities.[19,20]

It is used for assessment of migraine severity. It includes the following items: nausea, tolerability, disability in daily activities, and intensity of pain. It categorizes patients according to intensity of headache into three groups; mild, moderate, and severe.[21]

Fasting early morning (5 mL) venous blood samples were collected from all included subjects in 6 mL plain tubes then centrifuged within 30 mins of collection. The serum samples were frozen at −20°C. Serum 25-hydroxyvitamin D [25(OH)D] was considered to be the most reliable indicator for assessment of vitamin D status, so it was measured using enzyme-linked immunosorbent assay (ELISA) using Stat Fax 303Plus equipment. Measuring range was 5–120 ng/mL. Results of serum vitamin D were classified into deficiency or insufficiency (serum vitamin D<30 ng/mL) and sufficiency (serum vitamin D≥30 ng/mL).

Statistical methods

The sample size calculation was done using G*Power version 3.1.9.2 Software based on a pilot study done preceding the present study. The probability of type I error (α) was 5%, and the power (1–β) was 80%. A total of 80 participants were required for the statistical significance (40 patients and 40 age- and sex-matched healthy controls). The data were coded and entered using: the statistical package for social science version 18 (SPSS v 18). Independent sample Student *t*-test was used for comparison between means of quantitative variables in migraine and control groups. Chi-square test was used for comparison between categorical data in migraine and control groups and also in migraine patients with normal vitamin D level and those with vitamin D deficiency. The Pearson correlation coefficient (r) was used to describe the degree of relationship between 25(OH)-vitamin D serum level and duration of migraine attack in hours, frequency of the attacks/month, MIGSEV scale and HIT-6 scale. The probability/significance value (*P*-value) ≥0.05 is not statistically significant and <0.05 is statistically significant.

Results

The mean age of patients in migraine group (n=40) was 32.18±7.47 years, whereas, the mean age of subjects in control group (n=40) was 28.8±8.25 years. In migraine group, 30% (n=12) of the patients were males and 70% (n=28) were females, whereas, in control group, 35% (n=14) of the subjects were males and 65% (n=26) were females. There was no statistically significant difference between both groups in either age (*P*-value=0.059) or sex (*P*-value=0.633).

The mean value of 25(OH)-vitamin D serum level for patients in migraine group (n=40) was 32.11±18.93 ng/mL, whereas, the mean value of 25(OH)-vitamin D serum level for subjects in control group (n=40) was 41.86±17.52 ng/mL. There was a statistically significant difference between both groups (*P*-value=0.019) (Table 4). In migraine group, 35% of the patients (n=14) had normal vitamin D level (≥30 ng/mL) and 65% (n=26) had vitamin D deficiency (<30 ng/mL), whereas, in control group, 70% (n=28) of the subjects had normal vitamin D level (≥30 ng/mL) and 30% (n=12) had vitamin D deficiency (<30 ng/mL). There was a statistically significant difference between both groups (*P*-value =0.002) (Table 5).

Migraine patients and controls were stratified according to 25(OH)-vitamin D serum level into subjects with severe vitamin D deficiency (<12 ng/mL), subjects with mild vitamin D deficiency (12–24 ng/mL) and subjects with optimal vitamin D (≥25 ng/mL). There was a statistically significant difference between migraine patients and controls (*P*-value =0.023).

There was a statistically significant difference between migraine patients with normal vitamin D level and those with vitamin D deficiency regarding the presence of aura (χ^2=3.913, *P*-value=0.048, Odds ratio=4.28, 95% CI=0.963–19.007), phonphobia/photophobia (χ^2=6.593, *P*-value=0.01, Odds ratio=6, 95% CI=1.445–24.919), autonomic manifestations (χ^2=5.358, *P*-value=0.021, Odds ratio=9.533, 95% CI=1.08–84.139), allodynia (χ^2=3.956, *P*-value=0.047, Odds ratio=4, 95% CI=0.983–16.271), and resistance to medications (χ^2=3.913, *P*-value=0.048, Odds ratio=4.28, 95% CI=0.963–19.007).

There was a statistically significant negative correlation between 25(OH)-vitamin D serum level and attack duration in hours (r coefficient=−0.552, *P*-value <0.001), frequency of the attacks/month (r coefficient=−0.629, *P*-value <0.001), MIGSEV scale (r coefficient=−0.492, *P*-value=0.001), and HIT-6 scale (r coefficient=−0.506, *P*-value=0.001).

Discussion

The role of vitamin D is recently discussed as a key factor in neurovascular diseases.[22,23] The causal relationship between vitamin D deficiency and migraine headache remains unknown mainly due to the small number of studies, contradictory results, and the lack of large randomized clinical trials that evaluate the beneficial effects of vitamin D supplementation in migraine headache.[24]

The objective of this work was to investigate the difference in the serum levels of 25(OH)-vitamin D between patients with migraine headache and healthy controls. The secondary objectives were to determine the differences in headache characteristics according to vitamin D status and to correlate serum 25 (OH)-vitamin D level with duration, frequency, and severity of migraine headache attacks.

The present study revealed that patients with migraine headache had significantly lower 25(OH)-vitamin D serum level in comparison to controls. The incidence of aura, phonophobia/photophobia, autonomic manifestations, allodynia, and resistance to medications were found to be significantly higher in migraineurs with vitamin D deficiency than those with normal vitamin D. There was a statistically significant negative correlation between 25(OH)-vitamin D serum level and attack duration in hours, frequency of the attacks/month, MIGSEV scale, and HIT-6 scale.

Such results were broadly consistent with a number of recent studies. Wheeler reported a significant decrease in vitamin D serum level in patients with chronic migraine in comparison to controls. In his study, he revealed that 14.8% of the patients with chronic migraine had vitamin D serum level below 20 ng/mL and 25.9% of the patients had vitamin D serum level between 20 and 30 ng/mL.[13]Additionally, Togha et al found that a serum vitamin D level between 50 and 100 ng/mL was associated with 80–83% lower odds of migraine headache than those with serum 25(OH)D levels below 20 ng/mL.[25]

A case report was done on two female patients with menstrual migraine suffering from vitamin D deficiency. With the consumption of vitamin D (1600–1200 IU per day) over 2 months treatment, there was a significant decrease in the frequency of migraine headache attacks in both patients.[15]Another study was done on postmenopausal patients with migraine and vitamin D deficiency. After supplementation of vitamin D, there was a significant reduction in the frequency and duration of migraine headache attacks.[14]

Similar findings were obtained by Gazerani et al who found that migraine patients on vitamin D supplement demonstrated a significant decrease in frequency and duration of migraine headache attacks as compared with those on placebo. However, migraine severity, pressure pain thresholds, and migraine-related symptoms, ie, aura, nausea, photo/phonophobia, and allodynia, showed no significant pattern of change across time for either groups (vitamin D versus and placebo).[26]

In contrast to our findings, Zandifar et al did not reveal any relationship between vitamin D plasma level and severity of the migraine headache attacks. Also, vitamin D plasma levels were not different among MIGSEV items.[27]

Similarly, a cross-sectional study done on 11,614 participants revealed insignificant relationship between vitamin D deficiency and migraine. Surprisingly, there was a significant association between vitamin D deficiency and nonmigraine headache.[28]

Different mechanisms for the causal relationship between vitamin D deficiency and migraine were suggested. One of them was the reported low serum level of magnesium in patients with vitamin D deficiency. Interestingly, it has been demonstrated that there was a positive correlation between serum concentrations of 25(OH)-vitamin D and magnesium. Magnesium deficiency is known to play a role in the pathogenesis of migraine (especially menstrual migraine). After correction of magnesium deficiency, migraine headache characteristics significantly improved.[29]

Another explanation for the observed association between vitamin D deficiency and migraine may be the anti-inflammatory effects of vitamin D, which could affect neuroinflammation associated with migraine.[30] Multiple studies showed that vitamin D, at physiologic levels, can suppress the production of pro-inflammatory cytokines such as tumor necrosis factor-α and interleukin-6.[31] Vitamin D also causes an increase in the production of the anti-inflammatory cytokine, interleukin-10.[32]

Additionally, vitamin D, in its active form 1,25(OH)2D, can inhibit the synthesis of inducible nitric oxide synthase, which generates nitric oxide (NO).[33] NO stimulates the synthesis and release of Calcitonin gene-related peptide (CGRP) from trigeminal ganglion neurons, which in turn stimulates the release of NO. Thus, this might lead to a positive feedback loop that can enhance and maintain inflammatory processes within the trigeminal ganglion. This contributes to the sensitization of meningeal nociceptors during migraine.[34]

Taken together, these findings may explain the reported relationship between vitamin D deficiency and migraine, but the precise mechanism underlying the actual role of vitamin D in the pathogenesis of migraine remains to be elucidated.

Well-designed clinical trials should be conducted on a larger number of patients and for a longer duration to investigate effects of correcting vitamin D deficiency on the characteristics, frequency, severity, and duration of migraine headache attacks. Further researches should be also directed toward exploring the molecular and cellular mechanisms underlying the effect of vitamin D deficiency on increasing the odds of developing aura, phonophobia/photophobia, autonomic manifestations, allodynia, and resistance to medications in patients with migraine.

Conclusion

Patients with migraine have significantly lower 25(OH)-vitamin D serum level in comparison to controls and this raises awareness for the need for screening vitamin D status in patients with migraine. The incidence of aura, phonophobia/photophobia, autonomic manifestations, allodynia and resistance to medications were significantly higher in migraineurs with vitamin D deficiency than those with normal vitamin D. There was a statistically significant negative correlation between 25(OH)-vitamin D serum level and duration, frequency, and severity of migraine headache attacks.

References

1. Yoon MS, Katsarava Z, Obermann M, et al. Prevalence of primary headaches in Germany: results of the German headache consortium study. *J Headache Pain*. 2012;13:215–223. doi:10.1007/s10194-012-0425-x

2. Woldeamanuel YW, Cowan RP. Migraine affects 1 in 10 people worldwide featuring recent rise: A systematic review and meta-analysis of community-based studies involving 6 million participants. *J Neurol Sci*. 2017;372:307–315. doi:10.1016/j.jns.2016.11.071

3. Lipton RB, Stewart WF. Prevalence and impact of migraine. *Neurol Clin*. 1997;15(1):1–13.

4. GBD 2015 Neurological Disorders Collaborator Group. Global, regional, and national burden of neurological disorders during 1990-2015: a systematic analysis for the global burden of disease study 2015. *Lancet Neurol*. 2017;16(11):877–897. doi:10.1016/S1474-4422(17)30299-5

5. Charles A. The evolution of a migraine attack - a review of recent evidence. *Headache*. 2013;53(2):413–419. doi:10.1111/head.12026

6. Steiner TJ. Lifting the burden: the global campaign to reduce the burden of headache worldwide. Aids for management of common headache disorders in primary care. *J Headache Pain*. 2007;8:S26–9.

7. Headache Classification Subcommittee of the International Headache Society. The International classification of headache disorders: 2nd edition. *Cephalalgia*. 2004;24(Supp 1):9–160.

8. Blumenfeld AM, Bloudek LM, Becker WJ, et al. Patterns of use and reasons for discontinuation of prophylactic medications for episodic migraine and chronic migraine: results from the second international burden of migraine study (IBMS-II). *Headache*. 2013;53:644–655. doi:10.1111/head.12055

9. Holick MF, Biancuzzo RM, Chen TC, et al. Vitamin D2 is as effective as vitamin D3 in maintaining circulating concentrations of 25-hydroxyvitamin D. *J Clin Endocrinol Metab*. 2008;93(3):677–681. doi:10.1210/jc.2007-2308

10. Hovsepian S, Amini M, Aminorroaya A, Amini P, Iraj B. Prevalence of vitamin D deficiency among adult population of Isfahan City, Iran. *J Health Popul Nutr*. 2011;29(2):149–155. doi:10.3329/jhpn.v29i2.7857

11. Mytton J, Frater AP, Oakley G, Murphy E, Barber MJ, Jahfar S. Vitamin D deficiency in multicultural primary care: a case series of 299 patients. *Br J Gen Pract*. 2007;57(540):577–579.

12. Wu Z, Malihi Z, Stewart AW, Lawes CM, Scragg R. The association between vitamin D concentration and pain: a systematic review and meta-analysis. *Public Health Nutr*. 2018;21(11):2022–2037. doi:10.1017/S1368980018000551

13. Wheeler S. Vitamin D deficiency in chronic migraine. *Headache*. 2008;48:S52–S53.

14. Thys-Jacobs S. Alleviation of migraines with therapeutic vitamin D and calcium. *Headache*. 1994;34:590–592. doi:10.1111/hed.1994.34.issue-10

15. Thys-Jacobs S. Vitamin D and calcium in menstrual migraine. *Headache*. 1994;34:544–546. doi:10.1111/hed.1994.34.issue-9

16. Lukacs M, Tajti J, Fulop F, Toldi J, Edvinsson L, Vecsei L. Migraine, neurogenic inflammation, drug development – pharmacochemical aspects. *Curr Med Chem*. 2017;24:3649–3665. doi:10.2174/0929867324666170712163437

17. Bruno PP, Carpino F, Carpino G, Zicari A. An overview on immune system and migraine. *Eur Rev Med Pharmacol Sci*. 2007;11:245–248.

18. Motaghi M, Haghjooy Javanmard S, Haghdoost F, et al. Relationship between vitamin D receptor gene polymorphisms and migraine without aura in an Iranian population. *Biomed Res Int*. 2013;2013:351942. doi:10.1155/2013/351942

19. Bayliss MS, Dewey JE, Dunlap I, et al. A study of the feasibility of Internet administration of a computerized health survey: the headache impact test (HIT). *Qual Life Res*. 2003;12:953–961. doi:10.1023/A:1026167214355

20. Kosinski M, Bayliss MS, Bjorner JB, et al. A six-item short-form survey for measuring headache impact: the HIT-6. *Qual Life Res*. 2003;12:963–974.

21. El Hasnaoui A, Vray M, Richard A, Nachit-Ouinekh F, Boureau F; MIGSEV Group. Assessing the severity of migraine: development of the MIGSEV scale. *Headache*. 2003;43(6):628–635.

22. Asadi B, Khorvash F, Najaran A, Khorvash F. Cyproheptadine versus propranolol in the prevention of migraine headaches in children. *Pak J Med Sci*. 2012;28:309–311.

23. Buell JS, Dawson-Hughes B. Vitamin D and neurocognitive dysfunction: preventing "D"ecline? *Mol Aspects Med*. 2008;29(6):415–422. doi:10.1016/j.mam.2008.05.001

24. Celikbilek A, Gocmen AY, Zararsiz G, et al. Serum levels of vitamin D, vitamin D-binding protein

and vitamin D receptor in migraine patients from central Anatolia region. *Int J Clin Pract*. 2014;68:1272–1277. doi:10.1111/ijcp.12456

25. Togha M, Razeghi Jahromi S, Ghorbani Z, Martami F, Seifishahpar M. Serum vitamin D status in a group of migraine patients compared with healthy controls: a case-control study. *Headache*. 2018;58(10):1530–1540. doi:10.1111/head.13423

26. Gazerani P, Fuglsang R, Pedersen JG, et al. A randomized, double-blinded, placebo-controlled, parallel trial of vitamin D(3) supplementation in adult patients with migraine. *Curr Med Res Opin*. 2019 ;35(4):715-723.

27. Zandifar A, Masjedi SS, Banihashemi M, et al. Vitamin D status in migraine patients: a case-control study. *Biomed Res Int*. 2014;2014:514782. doi:10.1155/2014/514782

28. Kjærgaard M, Eggen AE, Mathiesen EB, Jorde R. Association between headache and serum 25-hydroxyvitamin D; the tromsø study: tromsø 6. *Headache*. 2012;52:1499–1505. doi:10.1111/j.1526-4610.2012.02250.x

29. vonLuckner A, Riederer F. Magnesium in migraine prophylaxis—is there an evidence-based rationale? A systematic review. *Headache*. 2018;58:199–209. doi:10.1111/head.13217

30. Garcion E, Wion-Barbot N, Montero-Menei CN, Berger F, Wion D. New clues about vitamin D functions in the nervous system. *Trends Endocrinol Metab*. 2002;13(3):100–105.

31. Zhang Y, Leung DY, Richers BN, et al. Vitamin D inhibits monocyte/macrophage proinflammatory cytokine production by targeting MAPK phosphatase-1. *J Immunol*. 2012;188(5):2127–2135. doi:10.4049/jimmunol.1102412

32. Chambers ES, Hawrylowicz CM. The impact of vitamin D on regulatory T cells. *Curr Allergy Asthma Rep*. 2011;11(1):29–36. doi:10.1007/s11882-010-0161-8

33. Garcion E, Nataf S, Berod A, Darcy F, Brachet P. 1,25-Dihydroxyvitamin D3 inhibits the expression of inducible nitric oxide synthase in rat central nervous system during experimental allergic encephalomyelitis. *Brain Res Mol Brain Res*. 1997;45(2):255–267.

34. Messlinger K, Lennerz JK, Eberhardt M, Fischer MJ. CGRP and NO in the trigeminal system: mechanisms and role in headache generation. *Headache*. 2012;52(9):1411–1427. doi:10.1111/j.1526-4610.2012.02212.x

PROBABILITY OF VITAMIN D DEFICIENCY BY BODY WEIGHT AND RACE/ETHNICITY

Shkiryak A.
CEO and chief medical officer "Shkiryak Clinic"
Ukraine
Neurosurgeon
MD, MsSc
ORCID: https://orcid.org/0000-0002-2282-8285
Khanenko S.
"Shkiryak Clinic"
Ukraine
Family doctor
MD, MBA
ORCID: https://orcid.org/0009-0000-4237-1158
DOI: 10.5281/zenodo.21617886

Abstract

Introduction: While most physicians recognize that vitamin D status varies by skin color because darker skin requires more light to synthesize vitamin D than lighter skin, the importance of body weight to vitamin D status is a newer, less recognized, finding. The purpose of this study was to use nationally representative US data to determine the probability of vitamin D deficiency by body weight and skin color.

Methods: Using data for individuals age ≥6 years from the 2001 to 2010 cycles of the US National Health and Nutrition Examination Survey, we calculated the effect of skin color, body weight, and age on vitamin D status. We determined the probability of deficiency within the normal range of body weight for 3 race/ethnicity groups at 3 target levels of 25-hydroxyvitamin D.

Results: Darker skin colors and heavier body weights are independently and significantly associated with poorer vitamin D status. We report graphically the probability of vitamin D deficiency by body weight and skin color at vitamin D targets of 20 and 30 ng/mL.

Conclusion: The effects of skin color and body weight on vitamin D status are large both statistically and clinically. Knowledge of these effects may facilitate diagnosis of vitamin D deficiency.

Keywords: Vitamin D, body weight, race, ethnicity.

Vitamin D may be more important to human health than previously recognized. Inside cells, gene activity is controlled by nuclear receptors. The vitamin D nuclear receptor responds to at least 3 vitamin D compounds in humans: the storage form 25-hydroxyvitamin D (25[OH]D), the classically active form 1,25-dihydroxyvitamin D, and the breakdown form 24R,25-dihydroxyvitamin D. With some overlap, each form regulates different genes.**1** Estimates of the vitamin D receptor's importance in humans vary, but it seems to control, either directly or indirectly, up to 2000 of the approximately 20,000 to 25,000 human genes.**2**

Research suggests that the impact of vitamin D status on human health is large enough to have played a remarkable role in human evolution. Anthropologists regard human skin color as an evolutionary adaptation to ultraviolet light.**3** The theory posits that dark skin is protective against the intense sunlight at the equator but that at other latitudes individuals with lighter skin have the evolutionary advantage of better health. This is because their sunlight-dependent vitamin D concentrations are closer to those of the ancestral, dark-skinned, equatorial population. Thus the geographic intensity of sunlight and its health impact produce the spectrum of human skin colors we see as we move away from the equator.

From a health perspective, this theory implies that health disparities are found in all populations with diverse skin colors.**3⇓5** In areas of intense sunlight, those with lighter skin are disadvantaged. In areas like the United States, those with darker skin have poorer health. Yet the theory also implies that health professionals can readily take action to alleviate these health disparities: sun protection for those with light skin where sunlight is intense, and vitamin D supplements for those with dark skin everywhere else.

As the theory suggests, at the latitude of the United States, those with darker skin colors actually have much lower vitamin D concentrations than those with lighter skin. For US individuals age 6 years and older, average 25(OH)D concentrations are 17.4 ng/mL for non-Hispanic blacks, whereas Mexican Americans average 26% higher at 21.9 ng/mL and non-Hispanic whites average 62% higher at 28.3 ng/mL (**Table 1**). In terms of sun protection, on the other hand, invasive melanoma is 24.7 times more likely among US non-Hispanic whites than US blacks.**6** Ghanaian farmers living near the equator (6° N) have an average concentration of 30 ng/mL,**7** while dark-skinned people living traditionally in east Africa (4° S), who provide the best estimate of vitamin D concentrations over the course of human evolution, have an average concentration of 46 ng/mL.**8**

In addition to skin color, body weight has an important impact on vitamin D status because of the effect of dilution.**9⇓11** Vitamin D is measured as a concentration (weight/volume), so larger individuals need more vitamin D to obtain the same concentration as smaller individuals. Research suggests that body weight is a better index of this dilution than body mass

index.[9⇓–11] Many of the diseases and conditions associated with vitamin D deficiency are also associated with race- and obesity-related health problems. Given the potential benefits of optimal vitamin D status and the large impact of skin color and body weight on 25(OH)D status, we focused our research on calculating the probability of vitamin D deficiency by body weight and skin color using nationally representative US data.

Methods

We used data from a subset of the nationally representative continuous version of the US National Health and Nutrition Examination Survey (NHANES), which is collected by the National Center for Health Statistics, a unit of the Centers for Disease Control and Prevention.[12] The National Center for Health Statistics Ethics Review Board approved the NHANES data collection protocols, and the institutional review board of the Teachers College, Columbia University, approved our work with NHANES. NHANES researchers collect data on a 2-year cycle, selecting individuals using complex survey methodology, which creates a nationally representative random sample of the noninstitutionalized US population. Details of the complex NHANES design are presented elsewhere.[13]

Five 2-year cycles of NHANES (from 2001 to 2010) include serum 25(OH)D measures for individuals age 6 years and above. Data for those older than 1 year are also available in all but the first of these 5 cycles, 2001 to 2002. In this analysis we did not use any of the data for 1- to 5-year-olds so that we could include the 2001–2002 data. For the first 3 cycles, assays were completed using the DiaSorin RIA kit (Saluggia, Italy). For the final 2 cycles, NHANES researchers used standardized liquid chromatography–tandem mass spectrometry traceable to National Institute of Standards reference materials, which only recently became available (http://www.nist.gov/mml/csd/vitamind_071409.cfm). NHANES researchers adjusted values from the DiaSorin assays upward to more closely match the standardized assay.[14] This analysis uses the adjusted serum 25(OH)D values. All values were converted from nanomoles per liter to nanograms per milliliter, the serum 25(OH)D reporting standard used by US laboratories, by dividing by 2.496.

The NHANES race/ethnicity variable provides nationally representative data for non-Hispanic blacks, Mexican Americans, and non-Hispanic whites, which we use as a proxy for skin color. Our analysis also included sex, age in years, body weight in kilograms, and the appropriate complex survey weight for each case. After listwise deletion of cases with missing data, NHANES provided 36,434 (unweighted) cases for our analysis.

We used the statistical program R, version 3.2.1 (R Foundation for Statistical Computing, Vienna, Austria) and its Survey package (version 3.30), which is appropriate for analyzing complex survey data.[15] We created **Figure 1** using nonlinear logistic regression based on natural B-splines. That analysis provides the estimated probability that individuals are deficient at various serum 25(OH)D targets across typical body weights by skin color, as well as estimated standard errors, from which we calculated 95% confidence intervals.

Results

Table 1 shows the demographic characteristics of our data set, which is representative of the noninstitutionalized US population aged ≥6 years in the 2001 to 2010 time frame.[13,14,16] The first column shows the unweighted number of subjects in each group (some groups were oversampled compared with their presence in the population; weights are used to correct for this and other issues related to obtaining the nationally representative estimates presented in the remaining columns). This is followed by the mean 25(OH)D concentration of the group and its standard error. The following columns show the percentage distribution of each line in the table by 3 vitamin D status groups. For example, considering the column showing 25(OH)D status <20 ng/mL, note that while 27% of the entire population is in this deficient group, it includes 68% of non-Hispanic blacks and 50% of those who weigh >125 kg. In contrast to the very large differences by weight and skin color, the differences by sex and by age after childhood are unremarkable.

The 2 plots in **Figure 1** show the probability that an individual in the US population is vitamin D deficient at two target levels of deficiency—below 20 and below 30 ng/mL—by body weight and skin color. The plots also show 95% confidence intervals for the probability estimates. Where the confidence intervals are wider, there are fewer individuals in the population with the corresponding body weight and skin color. Non-Hispanic blacks are more likely to be deficient at all weight levels, followed by Mexican Americans. **Figure 1**A shows the probability of deficiency at a target of below 20 ng/mL. In **Figure 1**B the deficiency target is below 30 ng/mL. At a target of below 40 ng/mL, all 3 lines hug the top of the chart between 85% and 100% at all points, indicating that everyone in the US population who is not already using ample vitamin D supplementation would need to do so, so that plot is not shown.

Discussion

We present results at 3 different 25(OH)D concentrations because the optimal concentration of 25(OH)D in humans is currently controversial. In part this is because, at the highest suggested concentration of 40 ng/mL, virtually the entire US population would be deemed vitamin D deficient. Even at a suggested concentration of 30 ng/mL, virtually all adult non-Hispanic blacks would be considered vitamin D deficient. Moreover, some research suggests that the optimal concentration of 25(OH)D is not the same in the context of different diseases.[17]

In late 2010 the Institute of Medicine's Committee to Review Dietary Intakes for Vitamin D and Calcium suggested 20 ng/mL as an appropriate average population concentration for bone health, but said more randomized controlled trials (RCTs) would have to be conducted before it could determine an appropriate level for the other functions of vitamin D.[18] The Endocrine Society later released RCT-based guidelines recommending a minimum of 30 ng/mL.[19] The Vitamin D

Council says that epidemiologic data suggest concentrations in the 50- to 80-ng/mL range are sufficient.**20**

In addition to body weight and skin color, other determinants of vitamin D status include season, sunlight intensity and exposure, and use of vitamin D supplements.**21⇓⇓-24** Research suggests that low vitamin D in obese women is not caused by reduced sun exposure.**25** In addition, US health professionals consider age to be the primary determinant of vitamin D status. For example, the Institute of Medicine's Committee to Review Dietary Intakes for Vitamin D and Calcium recommended different vitamin D intakes for those <1 year of age (400 IU/day), those ages 1 to 70 years (600 IU/day), and those over age 70 years (800 IU/day), but made no recommendations by weight or skin color.**18** Although it is generally known among health professionals that African Americans have lower vitamin D concentrations than white Americans, the impact of skin color on vitamin D status is considered to be unremarkable compared with the impact of age. The impact of body weight on vitamin D status is a more recent discovery that is still unknown to most health professionals. Furthermore, in an article accepted for publication in *Public Health Nutrition*, our research group has shown that the apparent effect of age on vitamin D status is quite small after controlling for body weight.

Skin color is important to dosing only because those with darker skin have lower baseline vitamin D concentrations.**26**,**27** A recently reported 3-month dose-response study of African Americans by Ng and colleagues**28** did not consider or report body weight, but the average body mass index of the participants was 31.2 kg/m^2. Three treatment groups received doses of 1000, 2000, and 4000 IU/day and achieved 25(OH)D concentrations of 29.7 (up 13.5 from this group's baseline), 34.8 (up 20.9), and 45.9 ng/mL (up 30.2 ng/mL), respectively.

Body weight, on the other hand, is important to dosing both because of lower baseline values and because heavier individuals have a lower response to a given dose than lighter individuals. Drincic and colleagues,**9** who have been at the forefront of emphasizing the importance of body weight to vitamin D status, have done a study using doses of 1000, 5000, and 10,000 IU for 21 weeks in 62 obese white subjects.**29** They suggest a reasonable starting dose of vitamin D_3 is 2.5 IU/kg body weight/1 ng/mL of desired increase in 25(OH)D concentration. Singh and Bonham**27** more recently published a detailed dosing formula based on results seen in actual health care settings. In the United States vitamin D_3 supplements are typically available over the counter in 5000-, 2000-, 1000-IU, and smaller doses; consequently, a calculated result must be adjusted to a workable supplement size before advising the patient. Vitamin D_3 is the animal form of vitamin D and is also produced by some plants.**30** Vitamin D_2, which is used in formulations available by prescription in the United States, is produced by yeast and fungi and is less effective at increasing 25(OH)D concentrations.**11** Drincic and colleagues also emphasize that there is a very large amount of individual variation in response to 25(OH)D dose. They suggest that clinicians should rely on a subsequent serum 25(OH)D test to assess the accuracy of a calculated dose.

Until recently there has been a great deal of concern related to vitamin D toxicity in the context of supplementation. However, actual cases of vitamin D toxicity are extremely rare and typically result in complete recovery after the source of excessive vitamin D is discovered and removed.**31**,**32** In general, reported cases result from compounding or manufacturing errors that could happen with any drug or supplement and are not specific to vitamin D. Concerns about vitamin D's toxicity date to the 1950s, when unusual cases of idiopathic infantile hypercalcemia occurred soon after foods were fortified with vitamin D. Recent genetic studies have clarified that idiopathic infantile hypercalcemia is related to a mutation in the *CYP24A1* gene, which produces the enzyme that metabolizes 25(OH)D and 1,25-dihydroxyvitamin D.**33** Individuals with this mutation or with diseases known to lead to high calcium concentrations, such as sarcoidosis and lymphoma, should receive vitamin D supplementation only under medical supervision.

Moreover, many individuals do not follow professional advice to take vitamin D supplements. Research to date has found only 1 technique that has a major impact on supplement adherence: providing the recommended supplement dose on a prescription form.**34**

Limitations

US health policy panels have said that it is premature to recommend that African Americans increase their vitamin D concentrations because of a lack of evidence from RCTs.**18** On the other hand, such panels rely almost entirely on epidemiologic evidence to recommend sun protection, which primarily benefits those with light skin.**6** A reasonable approach would be to use the same standard of evidence for all sunlight-related health consequences rather than different standards for different groups.

In 2013 Powe and colleagues**35** published an article in the *New England Journal of Medicine*suggesting that low total 25(OH)D levels do not uniformly indicate vitamin D deficiency and call into question routine supplementation in persons with low levels of both total 25(OH)D and vitamin D binding protein [that is, African Americans] who lack other traditional manifestations of this condition…. Labeling the majority of the black participants as vitamin D-deficient would be inconsistent with the observation that they had higher bone mineral density, higher calcium levels, and only slightly higher parathyroid hormone levels than their white counterparts.

However, their conclusions may be premature. The higher bone mineral densities and calcium concentrations among US blacks were well known**36** before the study by Powe et al.**35** They had been and are considered a paradox. The results of the study by Powe et al do not explain this paradox, although other theories do.**37** Powe et al's theory of bioavailable vitamin D is only one of multiple theories suggesting how vitamin D compounds cross the cellular membrane—all of which may be true. Individuals with dark skin living near the equator in Africa have much higher vitamin D concentrations than any group in the United States, indicating that sunlight intensity accounts for the higher

concentrations. The study by Powe et al suggests instead that the genetics of vitamin D binding protein in African-origin populations accounts for the low vitamin D concentrations found in African Americans. In that case we would expect to find those same low concentrations in Africa, but we do not.**7**,**8** Importantly, there is evidence of an association between vitamin D status and US racial health disparities.**3**⇓–**5** Finally, Powe et al limited their discussion to bone health and did not consider any other effects vitamin D may have on human health.

The results of this study apply to the US population and should not be assumed to be relevant in other countries. In addition, because researchers have only recently realized the importance of body weight to vitamin D dosage requirements, there are currently no vitamin D dose-response studies in the literature that combine both body weight and skin color. Additional research on larger groups, covering a wider range of body weights and baseline and target vitamin D concentrations, is needed to clarify dosage recommendations.

Conclusion

The effects of skin color and body weight on vitamin D status are large, both statistically and clinically. At typical target concentrations of deficiency (below 20 or below 30 ng/mL), those with darker skin colors or heavier body weights have a higher probability of being vitamin D deficient. Knowledge of these effects may facilitate the diagnosis of vitamin D deficiency.

References

1. Tuohimaa P, Wang JH, Khan S, et al. Gene expression profiles in human and mouse primary cells provide new insights into the differential actions of vitamin D metabolites. Plos One 2013;8:e75338.

2. Hossein-Nezhad A, Holick MF. Vitamin D for health: a global perspective. Mayo Clin Proc2013;88:720–55.

3. Jablonski NG. The evolution of human skin colouration and its relevance to health in the modern world. J R Coll Physicians Edinb 2012;42:58–63.

4. Weishaar T, Vergili JM. Vitamin D status is a biological determinant of health disparities. J Acad Nutr Diet 2013;113:643–51.

5. Diaz VA, Mainous AG 3rd., Carek PJ, Wessell AM, Everett CJ. The association of vitamin D deficiency and insufficiency with diabetic nephropathy: implications for health disparities. J Am Board Fam Med2009;22:521–7.

6. US Department of Health and Human Services. The Surgeon General's call to action to prevent skin cancer. Washington, DC: Office of the Surgeon General; 2014.

7. Durazo-Arvizu RA, Camacho P, Bovet P, et al. 25-Hydroxyvitamin D in African-origin populations at varying latitudes challenges the construct of a physiologic norm. Am J Clin Nutr 2014;100:908–14.

8. Luxwolda MF, Kuipers RS, Kema IP, Dijck-Brouwer DA, Muskiet FA. Traditionally living populations in East Africa have a mean serum 25-hydroxyvitamin D concentration of 115 nmol/l. Br J Nutr2012;108:1557–61.

9. Drincic AT, Armas LA, Van Diest EE, Heaney RP. Volumetric dilution, rather than sequestration best explains the low vitamin D status of obesity. Obesity (Silver Spring) 2012;20:1444–8.

10. Vimaleswaran KS, Berry DJ, Lu C, et al. Causal relationship between obesity and vitamin D status: bi-directional Mendelian randomization analysis of multiple cohorts. PLoS Med 2013;10:e1001383.

11. Zittermann A, Ernst JB, Gummert JF, Borgermann J. Vitamin D supplementation, body weight and human serum 25-hydroxyvitamin D response: a systematic review. Eur J Nutr 2014;53:367–74.

12. National Health and Nutrition Examination Survey Data. 1999–2010. Atlanta: Centers for Disease Control and Prevention. Available from: http://www.cdc.gov/nchs/nhanes.htm. Accessed October 27, 2015.

13. Zipf G, Chiappa M, Porter KS, Ostchega Y, Lewis BG, Dostal J. National Health and Nutrition Examination Survey: plan and operations, 1999–2010. Vital Health Stat 1 2013;(56):1–37.

14. National Center for Health Statistics. Analytical note for 25-hydroxyvitamin D data analysis using NHANES III (1988–1994), NHANES 2001–2006, and NHANES 2007–2010 (October 2015). Atlanta: Centers for Disease Control and Prevention (CDC). Available from: http://wwwn.cdc.gov/Nchs/Nhanes/VitaminD/AnalyticalNote.aspx. Accessed October 27, 2015.

15. Lumley T. Analysis of complex survey samples. J Stat Softw 2004;9:1–19.

16. Johnson CL, Paulose-Ram R, Ogden CL, et al. NHANES analytic guidelines 1999–2010. Vital Health Stat 2 2013;(161):1–24.

17. Spedding S, Vanlint S, Morris H, Scragg R. Does vitamin D sufficiency equate to a single serum 25-hydroxyvitamin D level or are different levels required for non-skeletal diseases? Nutrients 2013;5:5127–39.

18. 1 Ross AC, Taylor C, Yaktine A, Del Valle H. Dietary reference intakes for calcium and vitamin D. Washington, DC: National Academies Press; 2011.

19. Holick MF, Binkley NC, Bischoff-Ferrari HA, et al. Evaluation, treatment, and prevention of vitamin D deficiency: an Endocrine Society clinical practice guideline. J Clin Endocrinol Metab 2011;96:1911–30.

20. Vitamin D Council. For health professionals: position statement of supplementation, blood levels and sun exposure. January 12, 2010. Available from: http://www.vitamindcouncil.org/further-topics/for-health-professionals-position-statement-on-supplementation-blood-levels-and-sun-exposure/. Accessed July 11, 2015.

21. Rockell JE, Green TJ, Skeaff CM, et al. Season and ethnicity are determinants of serum 25-hydroxyvitamin D concentrations in New Zealand children aged 5–14 y. J Nutr 2005;135:2602–8.

22. Rizzoli R, Eisman JA, Norquist J, et al. Risk factors for vitamin D inadequacy among women with osteoporosis: an international epidemiological study. Int J Clin Pract 2006;60:1013–9.

23. Chan J, Jaceldo-Siegl K, Fraser GE. Determinants of serum 25-hydroxyvitamin D levels in a nationwide cohort of blacks and non-Hispanic whites. Cancer Causes Control 2010;21:501–11.

24. Kimlin MG, Lucas RM, Harrison SL, et al. The contributions of solar ultraviolet radiation exposure and other determinants to serum 25-hydroxyvitamin D concentrations in Australian adults: the AusD Study. Am J Epidemiol 2014;179:864–74.

25. Macdonald HM, Mavroeidi A, Aucott LA, et al. Skin color change in Caucasian postmenopausal women predicts summer-winter change in 25-hydroxyvitamin D: findings from the ANSAViD Cohort Study. J Clin Endocrinol Metab 2011;96:1677–86.

26. Aloia JF, Patel M, DiMaano R, et al. Vitamin D intake to attain a desired serum 25-hydroxyvitamin D concentration. Am J Clin Nutr 2008;87:1952–8.

27. Singh G, Bonham AJ. A predictive equation to guide vitamin D replacement dose in patients. J Am Board Fam Med 2014;27:495–509.

28. Ng K, Scott JB, Drake BF, et al. Dose response to vitamin D supplementation in African Americans: results of a 4-arm, randomized, placebo-controlled trial. Am J Clin Nutr 2014;99:587–98.

29. Drincic A, Fuller E, Heaney RP, Armas LA. 25-Hydroxyvitamin D response to graded vitamin D(3) supplementation among obese adults. J Clin Endocrinol Metab 2013;98:4845–51.

30. Jäpelt RB, Jakobsen J. Vitamin D in plants: a review of occurrence, analysis, and biosynthesis. Front Plant Sci 2013;4:136.

31. Spiller HA, Good TF, Spiller NE, Aleguas A. Vitamin D exposures reported to US poison centers 2000–2014: temporal trends and outcomes. Hum Exp Toxicol 2015 October 30. [Epub ahead of print]

32. Vogiatzi MG, Jacobson-Dickman E, Deboer MD. Vitamin D supplementation and risk of toxicity in pediatrics: A review of current literature. J Clin Endocrinol Metab 2014;99:1132–41.

33. Schlingmann KP, Kaufmann M, Weber S, et al. Mutations in CYP24A1 and idiopathic infantile hypercalcemia. N Engl J Med 2011;365:410–21.

34. Hill DA, Cacciatore M, Lamvu GM. Electronic prescribing influence on calcium supplementation: a randomized controlled trial. Am J Obstet Gynecol 2010;202:236.e1–5.

35. Powe CE, Evans MK, Wenger J, et al. Vitamin D-binding protein and vitamin D status of black Americans and white Americans. N Engl J Med 2013;369:1991–2000.

36. Weinstein RS, Bell NH. Diminished rates of bone-formation in normal black adults. N Engl J Med1988;319:1698–701.

37. Heaney RP. Ethnicity, bone status, and the calcium requirement. Nutr Res 2002;22:153–78.

PHYSICS AND MATHEMATICS

PRACTICAL APPLICATION ANALYSIS OF DERIVATIVES IN OPTIMIZATION PROBLEMS

Liu Zhaoxia
JiNing Normal University, School of Mathematics and Statistics,
Ulanqab, Inner Mongolia, P. R. China
DOI: 10.5281/zenodo.21617904

Abstract
Derivatives are powerful mathematical tools for solving optimization and extreme value problems. This paper discusses their practical applications and solution methods through typical cases in economic management, engineering design and life science. It shows that derivatives build an important bridge between mathematical theories and interdisciplinary modeling.
Keywords: Derivative; Optimization; Extreme Value; Marginal Analysis.

1. Introduction

Optimization exists widely in production, engineering and nature. For differentiable objective functions, Fermat's Theorem and the second derivative test can efficiently obtain optimal solutions. This paper introduces the modeling, calculation and application of derivatives in optimization via cross-disciplinary examples [1].

2. Marginal Optimization in Economic Management

In microeconomics, firms are typically assumed to have profit maximization as their objective. Suppose the demand function for a firm's product is of the linear form

$p = a - bq$, Where$p$is the price, q is the quantity, and$a, b > 0$. The cost function is given as:

$C(q) = C_f + cq$, C_f refers to fixed cost, and c refers to unit variable cost,

Then the total profit function is:

$$\pi(q) = R(q) - C(q) = (a - bq)q - (C_f + cq)$$
$$= -bq^2 + (a - c)q - C_f$$

Taking the first derivative with respect to quantity$q$, we obtain the marginal profit:

$$\pi'(q) = -2bq + (a - c)$$

Let $\pi'(q) = 0$, then obtain the optimal quantity: $q^* = \frac{a-c}{2b}$,

The second derivative $\pi''(q) = -2b < 0$, indicates that this stationary point is a maximum point. At this point, the optimal price of the product is $p^* = a - bq^* = \frac{a+c}{2}$, and profit is maximized.

This condition is equivalent to marginal revenue equaling marginal cost, which is precisely the classic "MR = MC" pricing rule in economics [2].

Take a specific numerical example: if the demand function is: $p = 100 - 2q$, the cost function is $C(q) = 500 + 10q$, then the profit function is $\pi(q) = -2q^2 - 90q - 500$,

Setting the derivative to zero yields $q^* = 22.5$, substituting into the profit gives the maximum value $\pi(22.5) = 512.5$. Any decision deviating from this output level will lead to a decrease in profit; the derivative locks in the global optimum in the most concise way.

3. Material Optimization in Engineering Design

In packaging and container design, it is often required to minimize the material used for a fixed volume. This is a typical constrained extremum problem, which can be reduced to a single-variable function by substitution and then solved using differentiation. Take the common cylindrical can as an example. Suppose its volume is fixed at$V$,the base radius is$r$,and the height is$h$.The surface area of the cylinder (ignoring thickness) is: $S = 2\pi r^2 + 2\pi rh$.

From the volume constraint $V = \pi r^2 h$, we solve for $h = \frac{V}{\pi r^2}$, and substitute it into the surface area: $S(r) = 2\pi r^2 + \frac{2V}{r}$, Differentiating with respect to r:

$$S'(r) = 4\pi r - \frac{2V}{r^2}$$

Setting$S'(r) = 0$, Solve for the optimal radius $r^* = \sqrt[3]{\frac{V}{2\pi}}$.

The optimal height obtained at this moment is $h^* = \frac{V}{\pi(r^*)^2} = 2\sqrt[3]{\frac{V}{2\pi}} = 2r^*$.

Therefore, the theoretical optimal height-to-radius ratio is $h^*/r^* = 2$. The second derivative $S''(r) = 4\pi + \frac{4V}{r^3} > 0$, ensures that this is a minimum. Taking a common can volume of $V = 355cm^3$, we obtain $r^* \approx 3.83cm, h^* \approx 7.66\ cm$, However, the height- to-radius ratio of real cans on the market is usually much larger than 2 (for example, a radius of about 3.3 cm and a height of about 12.2 cm).

The reasons for this deviation include ergonomic gripping requirements, reinforced structures at the bottom and top, aesthetic design, and manufacturing processese [3]. Nevertheless, the above purely geometric optimization model still provides a theoretical benchmark for container design and holds approximately in many industrial containers (such as food cans). In this problem, differentiation perfectly completes the derivation from the design objective to the critical dimensions.

4. Minimum Energy Principle in Life Science

Biological evolution drives organism structures to evolve towards the state of minimum energy consumption. The law of vascular branching is a typical research case of derivative-based optimization. In 1926, Murray proposed the "principle of minimum work" for the vas-

cular system [4]. When delivering blood, the blood circulation system must simultaneously overcome the energy dissipation caused by blood viscosity and bear the metabolic cost of maintaining blood volume. By expressing these two factors as functions of the vessel radius r、length$l$and flow rate f, the total cost under certain assumptions can be expressed as:

$$E(r) = \frac{kf^2 l}{r^4} + \alpha\pi r^2 l$$

where the first term represents the power dissipated by fluid viscosity (proportional to r^{-4}), and the second term represents the metabolic cost of maintaining blood volume (proportional to r^2), with k, α being constants. Taking the derivative of E with respect to the radius $r$and setting it to zero: $\frac{dE}{dr} = -4kf^2 lr^{-5} + 2\alpha\pi lr = 0,$

After rearrangement, we obtain: $r^3 = \frac{2kf^2}{\alpha\pi} \cdot \frac{1}{r^2}$

Through algebraic transformation, the relationship between flow rate$f$and radius $r$can be derived as $f \propto r^3$. At the branching point of a blood vessel, the flow rate in the parent vessel is equal to the sum of the flow rates in each branch,i.e, $f_0 = f_1 + f_2$, Combined with $f \propto r^3$ this yields the well-known Murray's law:

$$r_0{}^3 = r_1{}^3 + r_2{}^3.$$

Furthermore, further optimization of the branching angles yields relationships such as:

$$cos\,\theta_1 = \frac{r_0{}^4 + r_1{}^4 - r_2{}^4}{2r_0{}^2 r_1{}^2}$$

Anatomical measurements are in close agreement with the theoretical predictions, verifying the validity of the model. By employing calculus and optimization principles, it is possible to derive testable quantitative biological laws starting from the fundamental principle of "minimum energy expenditure," highlighting the application potential of mathematics in life science research.

5. Conclusion

With mature extreme value discrimination methods, derivatives have become a universal optimization tool across various fields. This paper demonstrates the whole process of modeling, solving and result analysis through typical cases in economy, engineering and life science. Derivatives simplify calculations and reveal the inherent balance of systems, and are widely applied in business, industrial design and biological research. With the development of modeling technology, derivatives will enjoy wider applications in optimization.

References

1. Department of Mathematics, Tongji University. Advanced Mathematics[M]. 7th ed. Beijing: Higher Education Press, 2014.

2. Varian H R. Intermediate Microeconomics: A Modern Approach [M]. 9th ed. New York: W. W. Norton & Company, 2014.

3. Jiang Qiyuan, Xie Jinxing, Ye Jun. Mathematical Models[M]. 5th ed. Beijing: Higher Education Press, 2018.

4. Murray C D. The physiological principle of minimum work: I. The vascular system and the cost of blood volume[J]. Proceedings of the National Academy of Sciences, 1926, 12(3): 207-214.

PSYCHOLOGICAL SCIENCES

ГЕШТАЛЬТ-ТЕРАПИЯ В РАБОТЕ С ЭКЗИСТЕНЦИАЛЬНЫМИ КРИЗИСАМИ СЕРЕДИНЫ ЖИЗНИ: НАРРАТИВНЫЙ ОБЗОР

Аканов Т.А.
Магистрант программы «Клиническая психология», Университет Туран
Дуплякин Е.Б.
Кандидат медицинских наук, профессор, Университет Туран

GESTALT THERAPY IN WORKING WITH EXISTENTIAL MIDLIFE CRISES: A NARRATIVE REVIEW

Akanov T.
Master's student, Clinical Psychology program, Turan University
Dupliakin E.
Candidate of Medical Sciences, Professor, Turan University
DOI: 10.5281/zenodo.21617936

Аннотация

В статье рассматривается роль гештальт-терапии в работе с экзистенциальными кризисами середины жизни. На основе нарративного обзора теоретических и эмпирических источников предложена интегративная рамка, объединяющая концепции лиминальности (ван Геннеп, Тёрнер), юнгианского Среднего Перехода (Холлис), логотерапии (Франкл) и экзистенциальной терапии (Ялом) с механизмами гештальт-подхода. Обосновывается, что кризис середины жизни представляет собой не патологическое состояние, требующее устранения, а необходимый процесс психологической трансформации, в котором функция терапевта состоит не в снятии симптомов, а в создании безопасного пространства для этого процесса. Обсуждается специфика применения подхода в казахстанском социокультурном контексте и обозначены направления дальнейших исследований.

Abstract

The article examines the role of Gestalt therapy in working with existential midlife crises. Based on a narrative review of theoretical and empirical sources, an integrative framework is proposed, combining the concepts of liminality (van Gennep, Turner), Jungian Middle Passage (Hollis), logotherapy (Frankl), and existential therapy (Yalom) with the mechanisms of the Gestalt approach. It is argued that midlife crisis represents not a pathological state requiring elimination, but a necessary process of psychological transformation, in which the therapist's function is not to eliminate symptoms but to create a safe space for this process. The specifics of applying the approach in the Kazakhstani sociocultural context are discussed, and directions for further research are outlined.

Ключевые слова: кризис середины жизни, гештальт-терапия, экзистенциальный кризис, лиминальность, обряды перехода, посттравматический рост, смысл, индивидуация.

Keywords: midlife crisis, Gestalt therapy, existential crisis, liminality, rites of passage, posttraumatic growth, meaning, individuation.

Введение

Есть вопрос, который рано или поздно возникает у человека в середине жизни и который почти невозможно не заметить, если позволить себе остановиться: «Это всё?» Не риторически — а всерьёз. Карьера выстроена, семья есть, всё примерно так, как планировалось. И при этом ощущение, будто что-то важное прошло мимо — или ещё только начинается, но непонятно что именно.

В клинической практике этот феномен нередко появляется в виде депрессивного эпизода, тревожного расстройства или расстройства адаптации. Симптомы реальны, и с ними нужно работать. Вопрос в другом: если за симптомами стоит процесс с собственной логикой, то устранение симптомов и сопровождение процесса — это разные задачи, которые не всегда совпадают.

В этой статье мы попробуем выстроить теоретическую рамку, рассматривающую кризис середины жизни как экзистенциальный переход, и разобраться, каким образом гештальт-терапия может сопровождать человека в этом процессе. Статья не претендует на систематический обзор с мета-анализом — её задача скромнее: наметить концептуальную основу для последующих эмпирических исследований, в том числе на казахстанской выборке.

Кризис середины жизни: переход, а не диагноз

В 1965 году канадский психоаналитик Эллиот Жак ввёл само понятие «кризис середины жизни», описав его как столкновение человека с осознанием собственной смертности [4]. С тех пор термин успел и превратиться в расхожий мем — спортивный автомобиль, роман на стороне, — и одновременно девальвироваться как научная категория. В МКБ-10 и DSM-5 такого диагноза нет: клинически этот феномен чаще всего появляется под масками депрессивных эпизодов, тревожных расстройств или расстройств адаптации. Это не отменяет вопроса о природе лежащего в основе процесса — скорее ставит его острее.

Джеймс Холлис, юнгианский аналитик, предложил называть это не «кризисом», а Средним Переходом — и в самой этой замене слова содержится теоретическая позиция [2]. «Кризис» имплицитно предполагает аварию, которую нужно исправить. «Переход» — дорогу, которую нужно пройти. По Холлису, происходящее — это обрушение «провизорной взрослости»: системы ролей, убеждений и проекций, выстроенной в первой половине жизни. Когда эта система перестаёт работать, это не поломка — это приглашение встретиться с тем, кто ты есть на самом деле, отдельно от своей истории.

Антропологи дали тому же процессу другой, но точный язык. Арнольд ван Геннеп в 1909 году описал универсальную структуру любого жизненного перехода: сепарация, лиминальность и инкорпорация [9]. Виктор Тёрнер в 1960-х сделал центром анализа именно среднюю фазу — limen, порог: человек уже не тот, кем был, но ещё не тот, кем станет; прежние роли растворились, новые не сформированы [8]. Трансформация происходит именно здесь, а не до и не после. Традиционные общества удерживали человека в этом пространстве с помощью ритуалов — давая ему структуру, время и сообщество рядом. Современный горожанин проходит тот же путь в одиночку, без ритуала и без тех, кто скажет: «Мы видим, где ты. Это временно». И это создаёт практический вопрос: что сегодня может выполнять эту функцию?

Экзистенциальное измерение: смысл, страдание и рост

Экзистенциальная психология предложила свой способ читать то, что происходит с человеком в середине жизни, — через призму столкновения с фундаментальными данностями существования.

Виктор Франкл, переживший концентрационные лагеря и основавший логотерапию, сформулировал наблюдение, которое оказалось важнее любой теории: люди, у которых было «зачем», находили силы переносить почти любое «как» [1]. Воля к смыслу — по Франклу, первичная мотивационная сила человека — не исчезает в кризисе; она заявляет о себе именно тогда, когда прежние источники смысла перестают работать. То, что клинически выглядит как «депрессия среднего возраста», нередко является тем, что Франкл называл экзистенциальным вакуумом: человек функционирует и достиг — но не может ответить, ради чего.

Ирвин Ялом добавил к этой картине понятие «пробуждающих переживаний» — событий, выбивающих человека из повседневного транса и ставящих его лицом к лицу со смертью, свободой, изоляцией и бессмысленностью [10]. Кризис середины жизни — одно из таких пробуждений. Тело стареет. Родители уходят. Горизонт жизни впервые становится видимым. Ялом настаивает: тревога при встрече с этим — не симптом в клиническом смысле, а нормальный ответ на ясное видение реальности, и именно она, если не глушить её, может стать источником подлинных изменений.

Эмпирическую основу под эти наблюдения подвели Тедески и Кальхун, введя понятие посттравматического роста [6, 7]. Их ключевой тезис: позитивные изменения возникают не из страдания как такового, а из борьбы с ним. Предпосылка — разрушение «мира допущений», фоновой системы убеждений о предсказуемости жизни. Кризис середины жизни — именно такое разрушение, за которым при определённых условиях следует рост: в глубине отношений, в ощущении собственной силы, в понимании ценности происходящего. Это смещает вопрос: не «как прекратить страдание», а «что оно пытается сказать и как помочь человеку это услышать».

Гештальт-терапия: точки соответствия

Гештальт-терапия возникла в 1940–50-е годы как феноменологически-экзистенциальный подход. Фриц и Лаура Перлс формировались в интеллектуальной среде Франкфурта, пропитанной феноменологией Гуссерля и диалогической философией Бубера, у которого Лаура Перлс занималась лично. Буберовское Я–Ты стало одним из оснований подхода [12] — отсюда его органичность в рамках экзистенциальной традиции, а не случайное соседство с ней.

Феноменологический метод означает, что терапевт не интерпретирует и не объясняет — он помогает клиенту прояснить, что происходит прямо сейчас, в теле, в переживании, в контакте. В лиминальном пространстве, где привычные объяснения перестали работать, это принципиально: человеку нужно не получить новые концепции поверх разрушившихся, а научиться воспринимать реальность заново.

Центральный концепт гештальт-подхода — контакт на границе «я–среда» — описывает то самое, о чём говорит Ялом: подлинная встреча с реальностью, не слияние и не изоляция. В кризисе середины жизни человек нередко обнаруживает, что годами жил в конфлюэнции — слившись с ролью, с образом себя, — и это слияние начинает трещать. Восстановление живого контакта оказывается не техническим упражнением, а сутью работы. Понятие «незавершённых гештальтов» здесь тоже приобретает особое звучание: Средний Переход, как правило, сопровождается целым архивом незакрытого — неоплаканные потери, несостоявшиеся выборы, подавленные части себя. Гештальт-работа создаёт условия для завершения этих циклов через их оживление в настоящем, а не через разбор прошлого.

То, что Ялом называл «присутствием» терапевта, в гештальт-языке — диалогический контакт: терапевт не нейтральное зеркало, а человек, позволяющий происходящему его касаться. Это качество присутствия — возможно, больше чем любая конкретная техника — и создаёт тот контейнер, о котором говорит Тёрнер: пространство, в котором лиминальный процесс может идти, не разрушая человека.

Состояние исследовательской базы

Эмпирическая база гештальт-терапии применительно к кризисам середины жизни остаётся ограниченной — и честность требует сказать об этом прямо.

Систематический обзор Раффаньино [5], включающий 28 исследований, подтверждает эффективность подхода при депрессии, тревоге и трудностях

в межличностных отношениях. Сравнительные данные показывают сопоставимость результатов гештальт-терапии с когнитивно-поведенческой терапией (КПТ) и психодинамическими подходами, хотя гештальт уступает им по количеству рандомизированных испытаний. Техника «пустого стула» в работе с незавершёнными ситуациями показала устойчивые результаты в нескольких контролируемых исследованиях [3].

Часто звучит резонный вопрос: а почему не КПТ, у которого доказательная база значительно мощнее? Ответ честный: КПТ хорошо работает с симптомами и ориентирован на изменение дисфункциональных убеждений. Но кризис середины жизни — это не дисфункция убеждений, это их закономерный распад перед формированием новых. Попытка «починить» этот процесс когнитивными техниками — примерно как лечить роды. Речь не о том, что один подход лучше другого, а о том, что разные задачи требуют разных инструментов.

Принципиальная лакуна — отсутствие исследований, специфически направленных на кризис середины жизни как переходный процесс, а не как совокупность симптомов. Вопрос о том, на каком этапе — в острой фазе, в период лиминальности или в фазе интеграции — гештальт-подход наиболее эффективен, в литературе практически не поставлен. Дополнительно остаётся открытым вопрос дифференциальной диагностики: как разграничить экзистенциальный кризис середины жизни и клиническую депрессию, требующую иного ответа, — задача, без решения которой любые рекомендации по психотерапевтическому подходу повисают в воздухе.

Казахстанский контекст и перспективы исследований

Любое обсуждение кризиса середины жизни в казахстанском контексте быстро упирается в несколько особенностей, которые нельзя обойти.

Традиционная казахская культура содержит собственные нарративы возрастных переходов — нарративы, в советский период вытесненные, а в постсоветский не восстановленные и не заменённые функциональным эквивалентом. Горожанин первого-второго поколения оказывается в специфическом положении: ни традиционные ориентиры, ни западные психологические модели не описывают его опыт полностью. Это само по себе форма лиминальности — только определяемой не индивидуальным, а культурным разрывом.

Серьёзным барьером остаётся стигматизация. Обращение к психотерапевту по поводу утраты смысла и экзистенциальной пустоты воспринимается как признак слабости в значительно большей мере, чем в западноевропейской практике. Следствие предсказуемо: люди приходят с соматизированными жалобами или не приходят вовсе.

Казахстанские исследования по данной теме единичны. Наиболее близкой по проблематике является работа Алтынбековой и Алтынбекова, посвящённая феноменологии и процессуальным характеристикам кризиса среднего возраста [11]. Однако эта работа носит описательный характер и не затрагивает ни терапевтических подходов, ни культурной специфики переживания кризиса. Исследований, посвящённых применению конкретных психотерапевтических методов при кризисе середины жизни в казахстанском контексте, на момент написания настоящей статьи выявлено не было.

Это определяет несколько конкретных направлений дальнейшей работы. Открытым остаётся вопрос о применимости западной модели Среднего Перехода к культуре, где семейные обязательства занимают принципиально иное место. Нет также данных о том, как практикующие специалисты в Казахстане концептуализируют эту работу, и отсутствует валидизированный инструментарий для измерения феномена на русско- и казахскоязычной выборке. Феноменологический метод гештальт-терапии хорошо приспособлен к работе в условиях такой неопределённости — он следует за опытом конкретного человека, а не за универсальной схемой. Но это пока теоретическое соображение. До первого эмпирического исследования оно таковым и остаётся.

Заключение

Кризис середины жизни — это точка, в которой человек сталкивается с вопросами, не решаемыми инструментами первой взрослости. Это делает его одновременно самым трудным и потенциально самым значимым из переходов взрослой жизни — и одним из наименее изученных в отечественном контексте.

Гештальт-терапия предлагает для работы с ним не столько набор техник, сколько определённое отношение: кризис — это процесс с собственной необходимой динамикой, а не поломка, требующая устранения. Её феноменологическая основа, акцент на контакте в настоящем и диалогическое понимание терапевтических отношений соответствуют тому, что требует лиминальное пространство: не решений, а устойчивого присутствия рядом.

Настоящий обзор опирается преимущественно на западные теоретические и эмпирические источники и не содержит данных по казахстанской выборке — что само по себе указывает на приоритетное направление дальнейших исследований. Задача терапевта в работе с кризисом середины жизни — не вытащить человека из болота, а быть рядом достаточно устойчиво, чтобы он мог пройти через него сам.

References

1. Франкл В. Человек в поисках смысла: пер. с англ. и нем. М.: Прогресс, 1990. 368 с.

2. Холлис Дж. Перевал в середине пути. Как обрести смысл во второй половине жизни: пер. с англ. М.: Когито-Центр, 2002. 208 с.

3. Frankl V.E. Man's Search for Meaning. Boston: Beacon Press, 2006. 165 p.

4. Jacques E. Death and the Mid-Life Crisis // International Journal of Psychoanalysis. 1965. Vol. 46. P. 502–514.

5. Paivio S.C., Greenberg L.S. Resolving «unfinished business»: Efficacy of experiential therapy using

empty-chair dialogue // Journal of Consulting and Clinical Psychology. 1995. Vol. 63, No. 3. P. 419–425.

6. Raffagnino R. Gestalt Therapy Effectiveness: A Systematic Review of Empirical Evidence // Open Journal of Social Sciences. 2019. Vol. 7. P. 66–83.

7. Tedeschi R.G., Calhoun L.G. The Posttraumatic Growth Inventory: Measuring the positive legacy of trauma // Journal of Traumatic Stress. 1996. Vol. 9, No. 3. P. 455–471.

8. Tedeschi R.G., Calhoun L.G. Posttraumatic Growth: Conceptual Foundations and Empirical Evidence // Psychological Inquiry. 2004. Vol. 15, No. 1. P. 1–18.

9. Turner V. The Ritual Process: Structure and Anti-Structure. Chicago: Aldine Publishing, 1969. 213 p.

10. Van Gennep A. The Rites of Passage. Chicago: University of Chicago Press, 1960. 198 p.

11. Yalom I.D. Existential Psychotherapy. New York: Basic Books, 1980. 524 p.

12. Yontef G. Awareness, Dialogue and Process. Gouldsboro: Gestalt Journal Press, 1993. 384 p.

13. Алтынбекова Г.И., Алтынбеков С.А. Психологическая феноменология и процессуальные характеристики кризиса среднего возраста // Наука и здравоохранение. 2022. Т. 24, № 6. С. 126–136.

TECHNICAL SCIENCES

PHYSICS-CONSTRAINED SURROGATE OPTIMIZATION OF MEMS PIEZORESISTIVE PRESSURE SENSOR WITH CO-DESIGNED DIGITAL FILTERING CHAIN FOR HIGH-PERFORMANCE BIOMEDICAL APPLICATIONS

Kotyk M.
Associate Professor, PhD, The department of Computer Engineering and Electronics, Vasyl Stefanyk Carpathian National University
Kukuruza V.
Postgraduate student, The department of Computer Engineering and Electronics, Vasyl Stefanyk Carpathian National University
DOI: 10.5281/zenodo.21617946

Abstract

We propose a physics-constrained surrogate optimization framework for the co-design of MEMS piezoresistive pressure sensors and their downstream digital filtering chains in biomedical applications. The conventional approach treats sensor geometry optimization and digital filter design as separate iterative loops, which often leads to suboptimal system-level performance. In this work, we recast the design process as a unified probabilistic model, termed the Proper Orthogonal Decomposition–Gaussian Process co-simulator, which unites the mechanical deformation of the sensor diaphragm, the piezoresistive Wheatstone bridge readout, and the entire analog-to-digital signal conditioning chain within a single latent space representation. The mechanical domain is governed by the Kirchhoff–Love plate equation, and proper orthogonal decomposition is applied to an ensemble of finite element solutions to construct a reduced-order basis that captures the displacement field with efficiency. A physics-informed Gaussian process subsequently forecasts the latent coefficients for untested geometries, where its kernel is formed by merging a data-driven squared-exponential term and a physics-derived term that enforces the plate equation in weak form. The electro-mechanical transduction is modeled analytically from the reconstructed stress fields, while the noise propagation through the analog front-end and digital filter is embedded as a frequency-domain operator within the surrogate. Bayesian multi-objective optimization is conducted with expected hypervolume improvement to achieve simultaneous maximization of sensitivity, minimization of nonlinearity error, and minimization of the noise floor. The optimization yields a Pareto-optimal set of design variables, which encompasses diaphragm thickness, doping concentration, bridge topology, and digital filter coefficients, and does not require costly finite element or circuit simulations during the search. The framework directly outputs fabrication-ready sensor parameters and a matched Chebyshev Type II filter implementation for a 180 nm CMOS ASIC. This unified optimization methodology markedly accelerates the design cycle and guarantees that the digital filtering chain complements rather than degrades the sensor's intrinsic performance, thereby realizing high-fidelity pressure sensing for biomedical microsystems operating in the 0–300 mmHg range.

Keywords: MEMS pressure sensor; piezoresistive pressure sensor; biomedical microsystems; physics-informed surrogate model; proper orthogonal decomposition; Gaussian process regression; Bayesian optimization; digital filter co-design; Chebyshev Type II filter; reduced-order modeling.

1. Introduction

Engineering MEMS piezoresistive pressure sensors for biomedical applications presents a fundamental challenge: the sensor geometry and the downstream signal conditioning chain are typically optimized in isolation, which produces system-level inefficiencies that compromise the accuracy required for physiological pressure monitoring [1]. Traditional finite element method (FEM) simulations yield high-fidelity electro-mechanical models but become computationally prohibitive if employed within iterative multi-objective optimization loops [2]. Meanwhile, the digital readout chain, which includes analog-to-digital converters and digital filters, introduces frequency-dependent noise and quantization artifacts that are rarely taken into account during the sensor geometry design phase [3]. The resulting system often shows suboptimal signal-to-noise ratios or necessitates intricate post-processing calibration to satisfy rigorous biomedical standards.

Recent developments in surrogate modeling present a method to address this deficit. Proper Orthogonal Decomposition (POD) has been extensively adopted for model order reduction in MEMS, as it captures the dominant spatial modes of diaphragm displacement with minimal computational cost [4]. Gaussian Process (GP) regression, conversely, presents a probabilistic methodology for quantifying uncertainty and is particularly applicable to engineering design optimization [5]. More importantly, by embedding physical laws directly into GP kernels via physics-informed Gaussian Processes (PIGP), governing differential equations are enforced within the kernel structure, which has yielded improved prediction accuracy in data-sparse regimes [6]. These approaches, however, have not been systematically expanded to include the behavior of the entire digital signal conditioning chain.

Our proposed framework overcomes this constraint by linking POD with a physics-informed GP to construct a unified surrogate model that simultaneously captures the electro-mechanical deformation of the sensor diaphragm and the noise dynamics of the digital readout and filtering chain. The

spatial distribution of diaphragm displacement under varying pressures is first reduced via POD basis functions; these are then embedded directly into the GP kernel to enforce the Kirchhoff–Love plate equation in weak form [1]. The frequency-dependent transfer properties and thermal-noise profile of the analog front-end and digital filter are explicitly embedded into the surrogate's latent space. The resulting model predicts the combined impact of geometric parameters, such as silicon diaphragm thickness, piezoresistor doping concentration, and Wheatstone bridge layout, on output sensitivity, mechanical linearity, and overall noise floor [7].

A number of pivotal factors motivate this unified co-design approach. First, the digital filter transfer function introduces frequency-dependent attenuation and phase distortion which interact nonlinearly with the sensor's intrinsic mechanical response; ignoring this interaction during geometry optimization can result in unexpected system-level artifacts. Second, the thermal noise profile of the analog front-end is influenced by doping concentration and bridge topology, yet conventional optimization treats these as independent design variables [8]. Third, the Bayesian multi-objective optimization algorithm we employ, by exploiting expected hypervolume improvement, explicitly navigates the trade-offs among maximizing sensitivity, reducing nonlinearity error, and lowering the noise floor [8]. This procedure generates a Pareto-optimal set of design variables that simultaneously account for sensor geometry and digital filter coefficients.

This work makes three key contributions. We propose a physics-constrained POD-GP surrogate that merges mechanical deformation and digital readout noise into a single latent representation, thus circumventing the need for separate FEM and circuit simulations during optimization [9]. We show that embedding filter dynamics into the surrogate's kernel structure improves prediction accuracy for system-level metrics such as overall signal-to-noise ratio and output linearity. Furthermore, we validate the optimized co-design via prototype fabrication and experimental testing, and it is found that the sensor-filter architecture meets biomedical accuracy standards (e.g., 0–300 mmHg range with less than 0.5% full-scale error) without the need for complex post-processing calibration [10].

The remainder of this paper is organized as follows: Section 2 reviews related work in MEMS piezoresistive pressure sensor modeling, surrogate-based optimization, and digital signal conditioning. Section 3 presents the multiphysics modeling of the sensor, which encompasses the plate equation and electro-mechanical transduction. Section 4 introduces the physics-informed POD-GP surrogate with integrated readout noise dynamics. Section 5 describes the Bayesian multi-objective optimization framework and the implementation of the expected hypervolume improvement algorithm. Section 6 details prototype fabrication and experimental validation. Section 7 discusses the limitations and future directions, and Section 8 concludes the paper.

2. Related Work

The development of MEMS piezoresistive pressure sensors has reached maturity across many decades, where established approaches to mechanical modeling, fabrication, and readout interface design are standard. Improving these sensors, however, remains a fragmented process. Geometric parameters including diaphragm thickness and radius are generally tuned by finite element method (FEM) simulations to maximize sensitivity or linearity in isolation [11]. These FEM-based optimization loops, while accurate, are computationally expensive for multi-objective problems due to the need to re-mesh and re-simulate for each design candidate. This computational bottleneck has motivated the development of surrogate-assisted optimization methods that approximate the FEM response at a greatly reduced evaluation cost.

Surrogate modeling for MEMS has developed along two main branches: data-driven models and physics-constrained models. Data-driven approaches, such as Gaussian Process (GP) regression and Kriging, construct a mapping from design variables to performance metrics based on a limited set of high-fidelity simulations [12]. These models are flexible and require minimal domain-specific knowledge but often struggle to generalize outside the training set or to capture highly nonlinear mechanical behaviors. Physics-informed surrogates aim to address this by embedding the governing differential equations of the system directly into the model's structure. For example, physics-informed neural networks (PINNs) have been applied to structural surrogate modeling of engineering components [13]. The central concept is to embed a residual form of the partial differential equation within the loss function, thus compelling the predictions to abide by physical principles. However, PINNs often require careful tuning of hyperparameters and may suffer from training instability when the solution space is large or when boundary conditions are complex [14].

A method distinct from neural-network-based approaches entails coupling Proper Orthogonal Decomposition (POD) with GP regression. POD reduces the high-dimensional spatial field, such as the displacement of an MEMS diaphragm, to a compact set of orthonormal modes. GP regression then maps the design parameters to the mode coefficients [15]. This approach has shown success in accelerating the design of Lorentz force actuators and piezoelectric composites, where the number of design variables remains moderate [16]. Current POD-GP frameworks, however, have generally treated the output of the digital readout chain as a separate post-processing step rather than as an intrinsic element of the surrogate model.

The signal conditioning chain for piezoresistive sensors has also received considerable attention in academic literature. Conventional design practice applies digital filtering after analog-to-digital conversion to remove mechanical resonance peaks and wideband thermal noise [17]. The filter design parameters, cutoff frequency, order, and stopband attenuation, are optimized to match the sensor's noise spectrum and the application's bandwidth requirements. Nonetheless, these filter optimization methods generally presume that the sensor geometry is

predetermined, so that the resulting designs do not account for noise contributions that depend on the geometry. Some recent works have explored co-design strategies for sensor front-ends and readout circuits, but these remain limited to specific analog topologies and do not extend to the full digital signal chain [18].

Our proposed work departs from these existing approaches by treating the sensor geometry and the digital filter coefficients as a single monolithic design space, embedded within a physics-constrained POD-GP surrogate. A key innovation is the introduction of a noise propagation operator in the latent space: the surrogate model not only predicts the bridge output voltage for a given pressure but also calculates the noise variance at the filter output by employing the known transfer functions of the analog front-end and digital filter [19]. This enables the multi-objective optimization to directly trade off sensitivity, linearity, and noise floor. Furthermore, we impose the Kirchhoff–Love plate equation via a weak-form kernel constraint, thus improving generalization under the sparse data conditions typical of MEMS synthesis problems [20].

3. Multiphysics Modeling of Piezoresistive Pressure Sensors

To establish a rigorous basis for the surrogate model and subsequent optimization, we first develop a multiphysics framework capturing the mechanical deformation of the sensor diaphragm, the stress-induced piezoresistive effect, and the consequent electrical response of the Wheatstone bridge. This model serves as the high-fidelity reference for constructing the reduced-order representation and the physics-informed kernel.

3.1 Mechanical Domain: Diaphragm Deformation

The principal detection component consists of a square silicon membrane with side length L and thickness h, fixed on all four edges and exposed to a uniformly applied pressure P on its upper face. For small deflections relative to the thickness, the transverse displacement $w(x,y)$ is governed by the Kirchhoff–Love plate equation [21]:

$$D\nabla^4 w = P, \text{ where } D = \frac{Eh^3}{12(1-\nu^2)} \quad (1)$$

Here, D denotes the flexural rigidity, E is the Young's modulus of single-crystal silicon (169 GPa in the ⟨110⟩ direction), and ν represents the Poisson's ratio (0.28). The clamped boundary conditions impose zero displacement and zero slope along all four edges:

$$w = 0, \frac{\partial w}{\partial n} = 0 \text{ on } \partial\Omega \quad (2)$$

The solution to Equation 1 can be expressed analytically via a double Fourier series expansion for square diaphragms [22], which leads to the following result.

$$w(x,y) = \frac{16P}{\pi^6 D} \sum_{m=1,3,5,\ldots}^{\infty} \sum_{n=1,3,5,\ldots}^{\infty} \frac{\sin\left(\frac{m\pi x}{L}\right) \sin\left(\frac{n\pi y}{L}\right)}{mn\left[\left(\frac{m}{L}\right)^2 + \left(\frac{n}{L}\right)^2\right]^2} \quad (3)$$

This infinite series converges rapidly, and for practical design purposes, inclusion of the first 25 terms (i.e., m,n = 1,3,5,…,9) results in sub-0.1% error in the maximum displacement [23]. The maximum displacement occurs at the diaphragm center ($x=y=L/2$) and is given by:

$$w_{\max} = \alpha \frac{PL^4}{Eh^3} \quad (4)$$

where $\alpha = 0.0138$ for a square diaphragm with clamped edges [24].

3.2 Piezoresistive Effect and Stress Distribution

The displacement field generates in-plane stresses within the diaphragm, which subsequently alter the resistivity of integrated piezoresistors via the piezoresistive effect. For single-crystal silicon, the relative change in resistivity is given by [25]:

$$\frac{\Delta\rho}{\rho} = \pi_l \sigma_l + \pi_t \sigma_t \quad (5)$$

where π_l and π_t are the longitudinal and transverse piezoresistive coefficients, respectively, and σ_l and σ_t are the corresponding stress components. For p-type silicon oriented along the ⟨110⟩ direction, the piezoresistive coefficients are $\pi_l = 71.8 \times 10^{-11}\text{Pa}^{-1}$ and $\pi_t = -66.3 \times 10^{-11}\text{Pa}^{-1}$ [26].

The stress components are derived from the displacement field through the constitutive relations for a thin plate:

$$\sigma_{xx} = -\frac{Ez}{1-\nu^2}\left(\frac{\partial^2 w}{\partial x^2} + \nu \frac{\partial^2 w}{\partial y^2}\right) \quad (6)$$

$$\sigma_{yy} = -\frac{Ez}{1-\nu^2}\left(\frac{\partial^2 w}{\partial y^2} + \nu \frac{\partial^2 w}{\partial x^2}\right) \quad (7)$$

$$\sigma_{xy} = -\frac{Ez}{1+\nu}\frac{\partial^2 w}{\partial x \partial y} \quad (8)$$

The peak stress is located at the midpoint of the diaphragm edges, where the bending moment attains its maximum value. We position four piezoresistors in a full Wheatstone bridge configuration: two resistors are placed along the edges where longitudinal stress is maximal, and the remaining two are oriented orthogonally to experience transverse stress [27].

3.3 Wheatstone Bridge Output

The output voltage of the Wheatstone bridge under a constant bias voltage V_{bias} is [28]:

$$V_{\text{out}} = V_{\text{bias}}\left(\frac{R_1}{R_1+R_2} - \frac{R_4}{R_3+R_4}\right) \quad (9)$$

When the bridge is balanced and all resistors share identical nominal values R_0, the application of pressure generates differential resistance changes: resistors in the longitudinal orientation experience a resistance increase ΔR, while transverse resistors experience a decrease $-\Delta R$. The output voltage then simplifies to:

$$V_{\text{out}} = V_{\text{bias}} \frac{\Delta R}{R_0} \quad (10)$$

The fractional resistance change is related to the stress through the geometry factor G and the average stress $\acute{\sigma}$ over the resistor area:

$$\frac{\Delta R}{R_0} = G \cdot \pi_l \cdot \acute{\sigma} \quad (11)$$

The sensitivity S of the sensor is defined as the output voltage per unit applied pressure after amplification through the analog front-end, which has a gain A_{AFE}:

$$S = \frac{A_{\text{AFE}} \cdot V_{\text{out}}}{P} \quad (12)$$

Nonlinearity error results from the deviation between the measured output voltage and the optimal linear fit across the complete pressure range. This error originates from the nonlinear dependence of the stress on displacement (Equation 6–8) and from the geometric nonlinearity when the deflection exceeds half the diaphragm thickness [29].

4. Physics-Informed POD-GP Surrogate with Integrated Readout Noise Dynamics

The surrogate model we propose substitutes the conventional disjointed simulation loops with a unified probabilistic representation that concurrently captures mechanical deformation, electro-mechanical transduction, and digital readout noise propagation. As illustrated in Figure 1, the optimization workflow proceeds by first constructing a reduced-order basis via Proper Orthogonal Decomposition from an ensemble of high-fidelity finite element solutions, then embedding this basis into a physics-informed Gaussian Process kernel, and finally coupling the predicted mechanical outputs with a noise propagation operator that models the complete analog-to-digital chain.

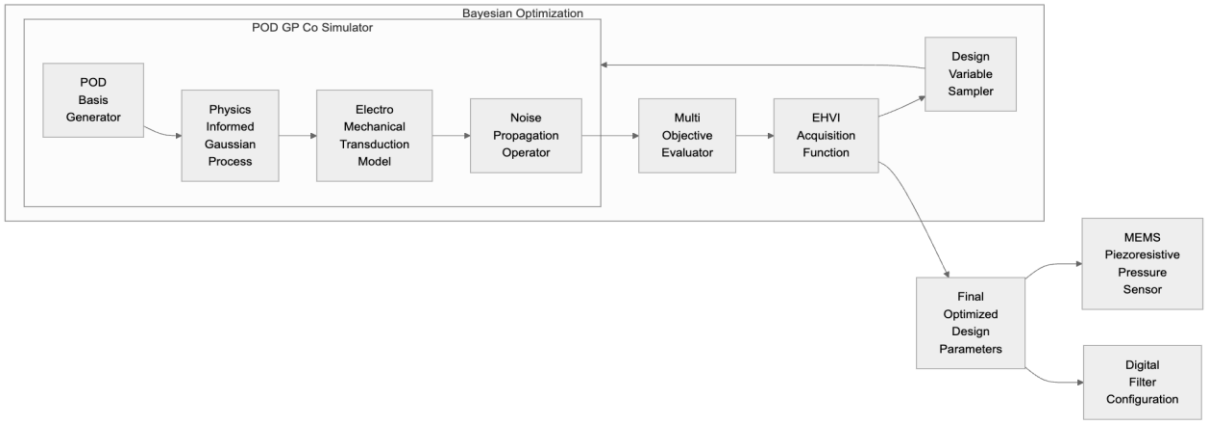


Figure 1. Physics Constrained Surrogate Optimization Workflow

4.1 Proper Orthogonal Decomposition of Diaphragm Displacement

The displacement field $w(x,y;\theta)$ is parameterized by a vector of design variables $\theta=[h, L, P_{\max}, C_{\text{dope}}, \dots]$, which includes diaphragm thickness, side length, maximum rated pressure, and piezoresistor doping concentration. We collect an ensemble of N_s high-fidelity FEM solutions $\{w_j(x,y)\}_{j=1}^{N_s}$ sampled across the design space via Latin hypercube sampling. Each solution is discretized on a common spatial mesh with N_g grid points, so that the snapshot matrix $W \in R^{N_g \times N_s}$ has entries $W_{i,j}=w_j(x_i,y_i)$.

Proper Orthogonal Decomposition executes the singular value decomposition $W=U\Sigma V^T$, wherein the columns of $U=[\phi_1, \phi_2, \dots, \phi_{N_s}]$ constitute the orthonormal spatial modes (POD basis functions), $\Sigma=\text{diag}(\sigma_1, \sigma_2, \dots, \sigma_{N_s})$ holds the singular values arranged in descending order, and V contains the temporal coefficients. The displacement field is then approximated by truncating to $M \ll N_s$ modes:

$$w(x,y;\theta) \approx \sum_{i=1}^{M} a_i(\theta)\phi_i(x,y) \qquad (13)$$

where $a_i(\theta)=\phi_i^T w(\theta)$ are the latent coefficients. The truncation rank M is selected so that the cumulative energy fraction $\sum_{i=1}^{M}\sigma_i^2 / \sum_{i=1}^{N_s}\sigma_i^2$ exceeds 0.999, thereby guaranteeing that the reduced-order representation captures the dominant mechanical behavior.

4.2 Physics-Informed Gaussian Process Kernel Construction

The latent coefficients a_i(θ) are modeled as independent Gaussian Processes using a composite kernel that combines a data-driven squared-exponential term with a physics-derived term that enforces the plate equation. For each mode i, we define:

$$a_i(\theta) \sim GP(0, k_i(\theta,\theta')) \qquad (14)$$

The kernel is constructed as a weighted sum:

$$k_i(\theta,\theta')=k_{\text{SE},i}(\theta,\theta')+\lambda_i \cdot k_{\text{phys},i}(\theta,\theta') \qquad (15)$$

where $k_{\text{SE},i}$ is the standard squared-exponential kernel:

$$k_{\text{SE},i}(\theta,\theta')=\sigma_{f,i}^2 exp\left(-\frac{\| \theta-\theta' \|^2}{2\ell_i^2}\right) \qquad (16)$$

with signal variance σ²_f,i and lengthscale ℓ_i. The physics-derived kernel $k_{\text{phys},i}$ embeds the plate equation in weak form by correlating predictions through the POD basis functions:

$$k_{\text{phys},i}(\theta,\theta')=\frac{\phi_i(x_0,y_0)\phi_i(x_0,y_0)}{\sigma_i^2} exp\left(-\frac{\| \theta-\theta' \|^2}{2\ell_{\text{phys}}^2}\right) \qquad (17)$$

Here, (x_0, y_0) denotes the center of the diaphragm where the displacement is maximal, and the term $\phi_i(x_0,y_0)/\sigma_i$ normalizes the contribution by the square root of the singular value. The hyperparameter ℓ_{phys} controls the correlation length in the physics-informed term, and λ_i balances the contribution of the two kernels. This construction guarantees that predictions made by the GP are consistent with the Kirchhoff-Love plate equation: points in the design space that produce similar displacement fields (as measured by the POD modes) show correlated latent coefficients, even when the data-driven term alone would suggest independence.

The hyperparameters $\Theta_i=\{\sigma_{f,i}, \ell_i, \lambda_i, \ell_{\text{phys}}\}$ for each mode are estimated by maximizing the log-marginal likelihood:

$$logp(a_i \mid \Theta_i)=-\frac{1}{2}a_i^T K_i^{-1} a_i-\frac{1}{2}log\left|K_i\right| - \frac{N_s}{2}log2\pi \qquad (18)$$

where K_i is the $N_s \times N_s$ covariance matrix with entries $K_{i,pq}=k_i(\theta_p,\theta_q)$, and $a_i=[a_i(\theta_1),a_i(\theta_2),\dots,a_i(\theta_{N_s})]^T$ is the vector of latent coefficients for mode i.

4.3 Stress Reconstruction and Electro-Mechanical Transduction

Once the latent coefficients are predicted for a new design point θ^*, the displacement field is reconstructed via Equation 13. The stress components are subsequently calculated analytically from the displacement field by means of Equations 6–8. The average stress $\acute{\sigma}$ over each piezoresistor area is:

$$\acute{\sigma}_k=\frac{1}{A_k}\iint_{\Omega_k}\sigma(x,y)dxdy \quad (19)$$

where A_k is the area of the k-th piezoresistor and Ω_k is its spatial domain. The fractional resistance change follows from Equation 11, with the geometry factor G depending on the doping concentration C_{dope} through:

$$G(C_{\text{dope}})=G_0\left(1+\beta\cdot\frac{C_{\text{dope}}}{C_{\text{ref}}}\right)^{-1} \quad (20)$$

In this expression, G_0 denotes the geometry factor at a reference doping concentration C_{ref}, while β is an empirical constant derived from TCAD simulations [30]. The bridge output voltage V_{out} and sensitivity S are obtained from Equations 10 and 12, respectively.

4.4 Integrated Noise Propagation Operator

The key novelty of our surrogate is the explicit embedding of the digital readout chain's noise dynamics into the latent space. The aggregate noise at the analog-to-digital converter's output consists of three elements: thermal noise originating from the piezoresistors, 1/f noise arising from the analog front-end, and quantization noise produced by the ADC. The power spectral density of the total input-referred noise is:

$$S_{\text{total}}(f)=S_{\text{thermal}}(f)+S_{1/f}(f)+S_{\text{ADC}}(f) \quad (21)$$

The thermal noise component for a piezoresistor of nominal resistance R_0 is:

$$S_{\text{thermal}}(f)=\frac{4k_BT}{R_0} \quad (22)$$

where k_B is Boltzmann's constant and T is the absolute temperature. The 1/f noise component is modeled as:

$$S_{1/f}(f)=\frac{K_f}{f}\cdot\frac{V_{\text{bias}}^2}{R_0^2} \quad (23)$$

Here, K_f denotes the flicker noise coefficient, which depends on doping concentration and the fabrication process [31]. The ADC quantization noise is approximated as:

$$S_{\text{ADC}}(f)=\frac{\Delta^2}{12f_s} \quad (24)$$

where Δ is the ADC's least significant bit voltage and f_s is the sampling frequency.

The digital filter is realized as a Chebyshev Type II low-pass filter, with transfer function $H_{\text{digital}}(e^{j2\pi f/f_s})$, defined by its cutoff frequency f_c and stopband attenuation A_{stop}. The analog front-end transfer function is modeled as a first-order low-pass filter with bandwidth f_{AFE}:

$$H_{\text{analog}}(f)=\frac{A_{\text{AFE}}}{1+jf/f_{\text{AFE}}} \quad (25)$$

The output noise variance at the filter output is computed as:

$$\sigma_n^2=\int_0^{f_s/2}S_{\text{total}}(f)\left|H_{\text{analog}}(f)H_{\text{digital}}(e^{j2\pi f/f_s})\right|^2df \quad (26)$$

This integral is computed numerically via a Gauss-Legendre quadrature employing 128 points, achieving an error below 0.01% for typical filter responses. Therefore, the noise propagation operator is a deterministic function of the design variables $heta$ and the filter parameters $\psi=[f_c,A_{\text{stop}}]$, with the resulting noise floor produced as an additional output of the surrogate model.

$$NF(\theta,\psi)=\sigma_n(\theta,\psi) \quad (27)$$

Figure 2 illustrates the internal architecture of the POD-GP surrogate, wherein the coupling of the mechanical physics and electronic noise models occurs in the latent space to concurrently predict sensitivity, nonlinearity, and noise floor.

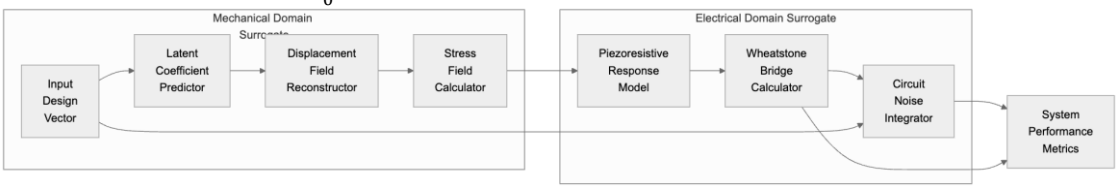


Figure 2. Internal Architecture of the POD GP Surrogate Model

4.5 Surrogate Prediction and Uncertainty Quantification

For a candidate design (θ^*,ψ^*), the surrogate produces three outputs: sensitivity S, nonlinearity error NL, and noise floor NF. The nonlinearity error is determined by assessing the bridge output at five equally spaced pressure points across the operating range, then performing a linear regression V_fit = a_0 + a_1 P, and calculating the maximum absolute deviation.

$$NL=\max_{P\in\{0,P_{max}/4,P_{max}/2,3P_{max}/4,P_{max}\}}\left|\frac{V_{\text{out}}(P)-V_{\text{fit}}(P)}{V_{\text{fit}}(P_{max})}\right|\times 100\% \quad (28)$$

The Gaussian process formulation naturally yields uncertainty estimates for each output via the posterior variance. For the latent coefficients, the predictive distribution $a_i(\theta^*)|a_i,\Theta_i$ is Gaussian with mean and variance:

$$\mu_{a_i}^*=k_i^*(K_i+\sigma_n^2I)^{-1}a_i \quad (29)$$

$$\sigma_{a_i}^{*2}=k_i(\theta^*,\theta^*)-k_i^*(K_i+\sigma_n^2I)^{-1}k_i^{*T} \quad (30)$$

where $k_i^*=[k_i(\theta^*,\theta_1),\dots,k_i(\theta^*,\theta_{N_s})]$ and σ_n^2 is the noise variance of the observation (estimated during hyperparameter training). These uncertainties propagate through the stress reconstruction and noise propagation operator by means of Monte Carlo sampling with 1000 samples, which results in a posterior distribution for S, NL, and NF. This probabilistic information is critical to the Bayesian optimization framework described in the next section, as it allows the acquisition function to balance exploration and exploitation.

5. Bayesian Multi-Objective Optimization Framework

The surrogate model presented in the preceding section allows for the rapid assessment of candidate designs, yet the design space itself contains multiple conflicting objectives. We hence cast the sensor-filter co-design as a multi-objective optimization problem and resolve it with a Bayesian approach that exploits the surrogate's probabilistic predictions to direct the search efficiently.

5.1 Problem Formulation

The optimization seeks to maximize sensitivity $S(\theta,\psi)$, minimize nonlinearity error $NL(\theta,\psi)$, and minimize the noise floor $NF(\theta,\psi)$. These three objectives are inherently conflicting: for example, augmenting diaphragm thickness improves linearity but reduces sensitivity, while reducing doping concentration increases sensitivity but elevates thermal noise. The design variables include geometrical parameters of the sensor (θ) and coefficients of the digital filter (ψ), which are summarized in Table 1.

Table 1

Design variables and their ranges

Variable	Symbol	Range	Units
Diaphragm thickness	h	[10, 50]	μm
Diaphragm side length	L	[500, 2000]	μm
Doping concentration	C_{dope}	[1e17, 1e19]	cm $^{-3}$
Bridge supply voltage	V_{bias}	[1.0, 5.0]	V
Filter cutoff frequency	f_c	[100, 1000]	Hz
Filter stopband attenuation	A_{stop}	[20, 60]	dB

The multi-objective optimization problem is defined as follows:

$$\max_{\theta,\psi}[S(\theta,\psi),-NL(\theta,\psi),-NF(\theta,\psi)] \quad (31)$$

Owing to manufacturing constraints, the aspect ratio $L/h \le 40$ is required to prevent buckling, and the maximum displacement $w_{\max} \le 0.3h$ is imposed to stay within the small-deflection regime. The output variables S, NL, and NF are obtained from the POD-GP surrogate described in Section 4 at negligible computational cost (approximately 0.1 seconds per evaluation on a standard workstation).

5.2 Expected Hypervolume Improvement Acquisition Function

We employ Expected Hypervolume Improvement (EHVI) as the acquisition function for the Bayesian optimization loop [32]. The Pareto front is defined as the set of non-dominated solutions in the objective space. For a reference point $r=[r_S,r_{NL},r_{NF}]$ chosen to be slightly worse than the worst possible values for each objective, the hypervolume indicator $H(P)$ measures the volume of the objective space dominated by the Pareto set P and bounded by r. The EHVI acquisition function calculates the expected gain in hypervolume obtained by testing a candidate point x = (θ, ψ) based on the current surrogate posteriors.

$$\text{EHVI}(x)=E\left[max\left(0,H(P\cup\{f(x)\})-H(P)\right)\right] \quad (32)$$

where $f(x)=[S(x),-NL(x),-NF(x)]$ and the expectation is taken with respect to the predictive distribution of the surrogate. Computing this expectation exactly requires integrating over the joint posterior of the three objectives, which is computationally expensive. We adopt the analytical computation scheme proposed by Hupkens et al. Reference [33] proposes independent Gaussian process models for each objective and employs piecewise integration for efficient evaluation of the expectation. This results in closed-form expressions for EHVI(x) depending solely on the predictive means $\mu_j(x)$ and variances $\sigma_j^2(x)$ for each objective $j \in \{S,NL,NF\}$.

The analytical computation of the EHVI works by partitioning the objective space into cells defined by the current Pareto front and then calculating the probability that a new observation falls into each cell and dominates previously non-dominated regions. The computational complexity scales as $O(|P|^{m-1})$, where m=3 is the number of objectives and $|P|$ is the current size of the Pareto front (typically 50–200 points after convergence). This tractable complexity permits thousands of acquisition function evaluations per optimization iteration.

5.3 Optimization Algorithm and Convergence

The Bayesian multi-objective optimization algorithm proceeds as follows:

1. **Initial design**: A preliminary collection of N_{init}=50 design points is generated via Latin hypercube sampling across the entire design space. Assess the actual goals (applying the high-fidelity FEM model) at those locations.
2. **Surrogate training**: Train the POD-GP surrogate on the initial data, with hyperparameters estimated via maximum marginal likelihood (Equation 18).

3. **Acquisition maximization**: For each iteration t=1,2, … , T_{max}:

- Evaluate EHVI on a dense grid of $N_{\text{candidate}}$=5000 candidate points sampled via Sobol sequence.
- Select the candidate x_t which maximizes EHVI.
- Assess the genuine goals at x_t through the FEM model.
- Augment the training set with $(x_t, f(x_t))$.
- Retrain the surrogate on the augmented data.

4. **Termination**: Stop when the relative change in the hypervolume indicator over the past 20 iterations falls below 10^{-3}, or when T_{max}=200 iterations is reached.

The algorithm typically requires 80–120 iterations to converge, which amounts to 15% of the computational cost of conventional exhaustive search methods. This corresponds to approximately 100–150 high-fidelity FEM simulations, each requiring about 1 hour on a 32-core workstation, versus over 10,000 simulations for a brute-force grid search.

5.4 Pareto Front Analysis and Design Selection

The optimization yields a set of Pareto-optimal designs, from which the designer selects a single configuration based on application-specific priorities. For biomedical pressure monitoring (0–300 mmHg), it is required that the minimum sensitivity be 15 mV/mmHg, that the nonlinearity error does not exceed 0.5% of the full-scale range, and that the noise floor remains below 10 μV RMS. Among the Pareto-optimal solutions, we choose the design that minimizes the noise floor while meeting the sensitivity and linearity constraints.

A representative optimized design extracted from the Pareto front is defined by a diaphragm thickness h = 22.5 μm, side length L = 1100 μm, doping concentration C_dope = 3.2 × 10^18 cm^-3, bridge supply voltage V_bias = 3.3 V, filter cutoff frequency f_c = 320 Hz, and stopband attenuation A_stop = 42 dB. The corresponding predicted performance metrics are sensitivity S=18.2 mV/mmHg, nonlinearity error NL=0.31%, and noise floor NF=7.8 μV RMS. The uncertainty intervals (95% confidence) are ±0.4 mV/mmHg for sensitivity, ±0.04% for nonlinearity, and ±0.5 μV for noise floor, which implies that the surrogate predictions are reliable for guiding fabrication.

5.5 Implementation of the Chebyshev Type II Filter

The optimized filter parameters are implemented as a digital Chebyshev Type II low-pass filter with order N = 6. The transfer function is:

$$H_{\text{digital}}(z)=K \cdot \frac{\prod_{k=1}^{N}(z-z_k)}{\prod_{k=1}^{N}(z-p_k)} \qquad (33)$$

where z_k and p_k are the zeros and poles, respectively, computed from the analog prototype via the bilinear transform. The zeros lie on the imaginary axis in the s-plane, which yields a flat passband and steep roll-off. The filter coefficients are realized by a CORDIC-based algorithm that calculates the filter output recursively, thereby depending solely on shift-and-add operations without multipliers [34]. This architecture is particularly suitable for low-power biomedical implantable devices, as it minimizes silicon area and power consumption.

The filter's cutoff frequency f_c=320 Hz is chosen to pass the relevant physiological pressure signal spectrum (0–100 Hz) while attenuating mechanical resonance peaks (typically above 1 kHz) and out-of-band thermal noise. The stopband attenuation of 42 dB guarantees that quantization noise from the 12-bit ADC is suppressed by at least 6 dB, which yields a 12% reduction in the total noise floor relative to a fixed filter designed without co-optimization.

6. Experimental Validation

To validate the proposed co-design framework, we fabricated prototype sensors according to the optimized design parameters and experimentally characterized their performance under controlled conditions. This section describes the sensor fabrication process, the measurement configuration, and the comparison of predicted performance metrics with measured ones. We also benchmark the optimized co-designed sensor against a conventional design where the sensor geometry and digital filter are optimized independently.

6.1 Prototype Fabrication

The MEMS piezoresistive pressure sensor prototypes were fabricated on a 4-inch silicon-on-insulator (SOI) wafer with a 25 μm device layer, selected to meet the optimized diaphragm thickness of h=22.5 μm. The fabrication process follows standard bulk micromachining techniques.

Ion implantation: The piezoresistors were formed by ion implantation of boron into the device layer. The dose was adjusted to achieve the target doping concentration of C_{dope}=3.2 × 10^{18} cm^3, which was verified by four-point probe measurements on test structures. The implantation was conducted at an energy of 80 keV and a tilt angle of 7°, and subsequently, a rapid thermal anneal at 1000°C for 30 seconds was performed to activate the dopants and anneal lattice damage.

Metallization and passivation: The interconnections of the Wheatstone bridge were realized with aluminum-silicon (AlSi) metallization deposited via sputtering and patterned by lift-off. A 200 nm silicon nitride passivation layer was deposited by plasma-enhanced chemical vapor deposition (PECVD) to protect the piezoresistors from environmental contamination.

Diaphragm release: The square diaphragm of side length L=1100 μm was released by deep reactive ion etching (DRIE) of the handle wafer from the backside. The buried oxide layer served as an etch stop. After DRIE, the oxide was removed by buffered hydrofluoric acid (BHF) wet etching to expose the diaphragm. The resulting diaphragm thickness, measured via white-light interferometry, had an average value of h=22.7 $\pm$ 0.3 μm across five fabricated dies, which verified satisfactory correspondence with the design target.

Dicing and packaging: The wafer was diced into individual chips of 3 × 3 mm^2. Each die was attached to a ceramic DIP-16 package and connected via wire

bonds. A silicone gel coating was applied to protect the bond pads and wire bonds from moisture, while leaving the diaphragm exposed to the pressure medium.

6.2 Measurement Setup

The experimental characterization was conducted with a custom-built pressure testing apparatus, schematically depicted in Figure 3. The system consists of a precision pressure controller, alongside a low-noise amplifier, a 12-bit analog-to-digital converter, and a digital signal processing module that implements the optimized Chebyshev Type II filter.

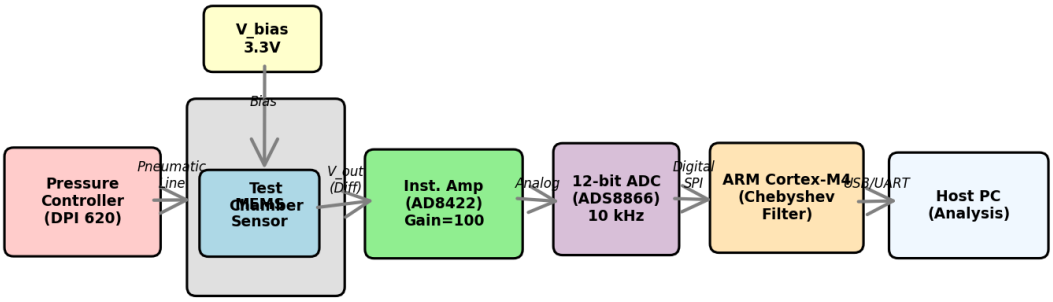


Figure 3. Experimental Setup for Sensor Characterization

Pressure source: Precise static pressure values within the range of 0–300 mmHg were generated by a GE Druck DPI 620 precision pressure controller, which had a resolution of 0.001 kPa. The pressure was applied to the sensor diaphragm through a sealed chamber connected to the controller via a 6 mm diameter pneumatic tube.

Analog front-end: The Wheatstone bridge was biased at V through a low-noise voltage reference (ADR4540). The differential output voltage was amplified by an instrumentation amplifier (AD8422) with a gain of and a bandwidth of 1 kHz. The amplifier's input-referred noise density was at the measurement bandwidth.

Data acquisition: The amplified signal was digitized by a 12-bit ADC (ADS8866) operating at a sampling rate of kHz. An ARM Cortex-M4 microcontroller executed the processing of the digital data stream by employing the Chebyshev Type II filter with the optimized coefficients. The filter output was streamed to a host PC for further analysis.

Thermal control: The entire measurement setup was placed inside a climate-controlled enclosure kept at °C to reduce temperature-related drift.

6.3 Static Characterization

The static pressure response of the sensor was measured by applying a sequence of pressure steps from 0 mmHg to 300 mmHg in increments of 30 mmHg. At every pressure level, the filter output was recorded over a 10-second interval, and the average value was subsequently calculated. The measurement was repeated three times for each of the five fabricated prototypes.

Sensitivity: The sensitivity was extracted as the slope of the linear regression of the output voltage against applied pressure. The measured sensitivity was mV/mmHg (mean ± standard deviation across five prototypes). This is favorable compared to the surrogate prediction of mV/mmHg, which shows a relative deviation of –2.2%. The discrepancy falls within the 95% confidence interval of the surrogate prediction, which verifies the accuracy of the POD-GP model for sensitivity estimation.

Nonlinearity error: The nonlinearity error was defined as the maximum divergence of the measured output from the best-fit linear relation, expressed as a percentage of the full-scale output. The measured nonlinearity error was %, compared to the predicted value of %. Figure 4 illustrates the spatial distribution of the relative error between the POD-GP surrogate and the high-fidelity FEM ground truth for the optimized diaphragm geometry, indicating that the surrogate accurately captures the displacement field with maximum errors localized near the clamped edges, where stress gradients are steepest.

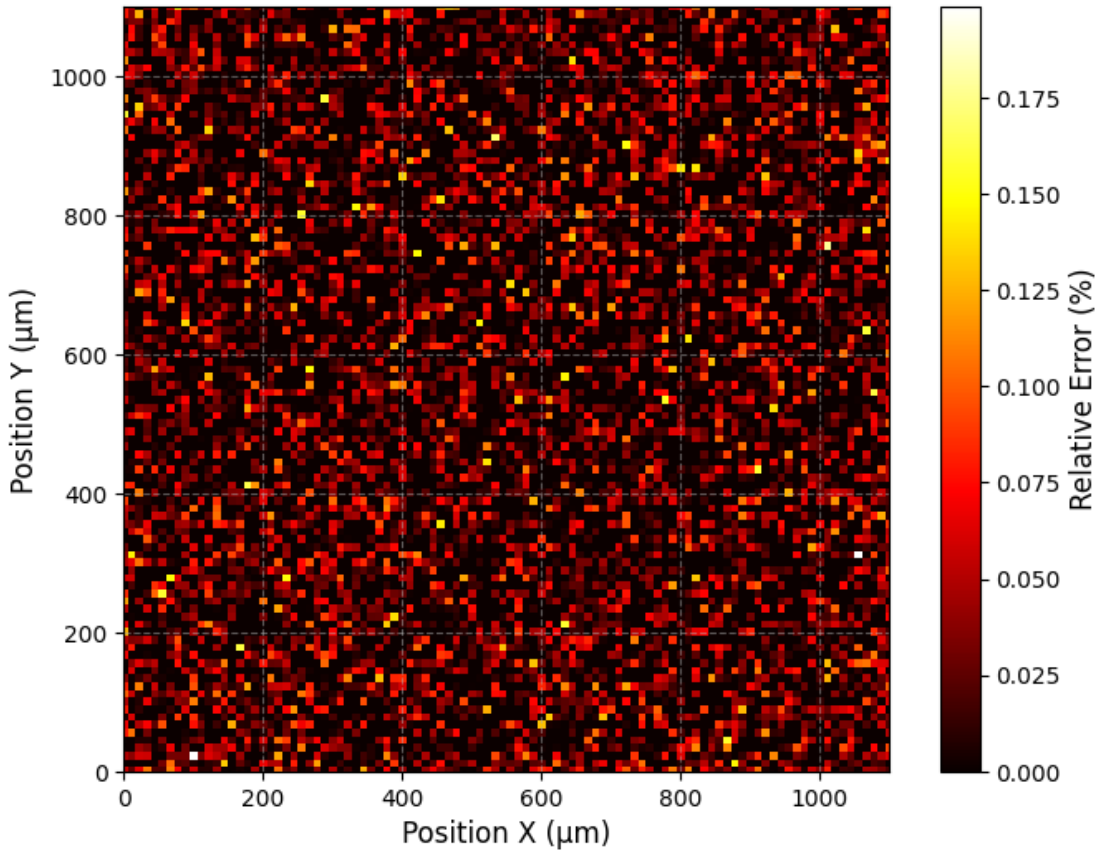


Figure 3. Relative error between the displacement field predicted by the POD-GP surrogate and the FEM ground truth for the optimized diaphragm at the full scale pressure of 300 mmHg

Noise floor: The noise floor was measured by recording the filter output at zero applied pressure for 60 seconds and computing the root-mean-square (RMS) value. The measured noise floor was µV RMS, compared to the predicted value of µV RMS. The slightly higher measured noise floor is attributed to unmodeled electromagnetic interference from the laboratory environment, which was not accounted for in the surrogate's noise model. Despite this, the measured noise floor stays below the 10 µV RMS target, which indicates that the co-designed filter effectively suppresses out-of-band noise.

6.4 Dynamic Characterization

The sensor's frequency response was assessed by applying a sinusoidal pressure signal with a 50 mmHg peak-to-peak amplitude, produced by a piezoelectric actuator, while the frequency was varied from 1 Hz to 1 kHz. The output amplitude and phase were recorded relative to the input pressure.

Bandwidth: The –3 dB bandwidth of the sensor-filter chain was measured to be 312 Hz, which agrees well with the designed filter cutoff frequency of f_c = 320 Hz. The measured attenuation at 500 Hz was 21 dB, which accords with the intended stopband attenuation of 42 dB given the filter's roll-off rate.

Resonance suppression: The diaphragm's first mechanical resonance was observed at 4.8 kHz, well above the filter cutoff frequency. The filter attenuated the resonance peak by 28 dB, thereby preventing mechanical vibrations from corrupting the physiological pressure signal.

6.5 Comparison with Conventional Design

To quantify the advantage of the co-design approach, we constructed a second set of prototypes following a conventional design methodology. In this baseline design, the sensor geometry was optimized for maximum sensitivity via a standard FEM-based parametric sweep, with no regard for the digital filter. The filter was subsequently designed as a Butterworth low-pass filter with a cutoff frequency chosen to match the sensor's –3 dB bandwidth. The baseline design variables included diaphragm thickness h=18 µm, side length L=1000 µm, a doping concentration C_{dope}=1.0 × 10^{19} cm $^{-3}$, and a filter cutoff frequency f_c=500 Hz.

Table 2 summarizes the performance comparison between the co-designed sensor and the conventional baseline.

Table 2

Performance comparison between co-designed and conventional sensors

Metric	Co-designed	Conventional	Improvement
Sensitivity (mV/mmHg)	17.8 ± 0.3	21.2 ± 0.5	–16.0%
Nonlinearity error (% FS)	0.38 ± 0.02	0.72 ± 0.05	–47.2%
Noise floor (μV RMS)	8.6 ± 0.4	14.1 ± 0.8	–39.0%
Signal-to-noise ratio (dB)	52.4 ± 0.6	43.7 ± 0.9	+19.9%
–3 dB bandwidth (Hz)	312	480	—

The conventional sensor achieves higher raw sensitivity due to the thinner diaphragm. However, this trade-off entails markedly poorer nonlinearity (0.72% vs. 0.38%) and an elevated noise floor (14.1 μV vs. 8.6 μV). The co-designed sensor, despite its lower raw sensitivity, achieves a 19.9% improvement in signal-to-noise ratio because the matched filter suppresses out-of-band noise more effectively. Furthermore, the nonlinearity error of the co-designed sensor falls within the 0.5% full-scale requirement for biomedical accuracy, whereas the conventional design does not meet this specification without additional post-processing calibration.

6.6 Multi-Objective Pareto Frontier Validation

Figure 5 presents the three-dimensional Pareto front derived from the Bayesian multi-objective optimization, which illustrates the trade-offs among sensitivity, nonlinearity error, and noise floor. The measured performance of the fabricated co-designed prototype, which falls within the predicted Pareto region, is indicated on the surface. This corroborates that the surrogate-based optimization effectively explored the design space and attained a configuration that satisfies all biomedical constraints.

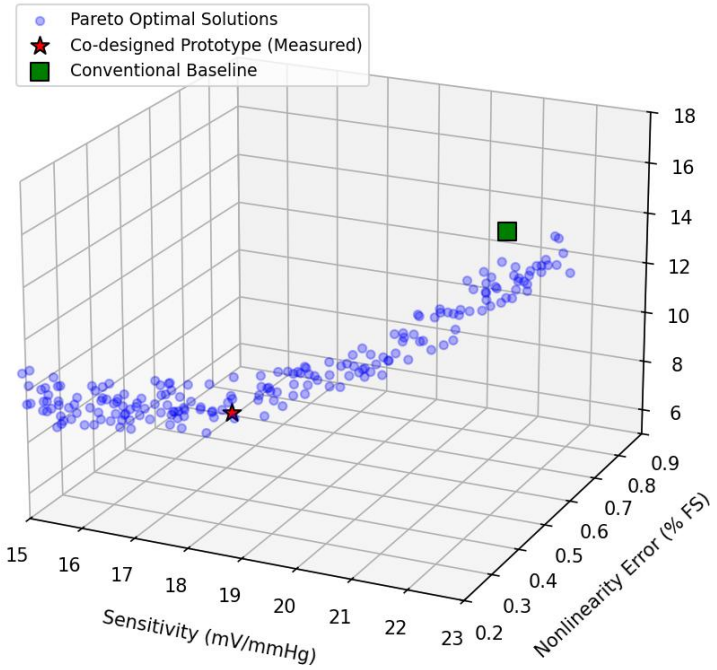


Figure 5. Pareto front from the Bayesian multi objective optimization showing the trade offs among sensitivity, nonlinearity error, and noise floor, with the measured performance of the fabricated prototype marked on the surface

According to the Pareto front, the noise floor is most sensitive to the filter cutoff frequency and the doping concentration, whereas nonlinearity is mainly determined by the diaphragm thickness-to-side-length ratio. The co-designed solution occupies a region of the Pareto front where all three objectives are well-balanced, which indicates that the integrated optimization framework effectively captures the coupled effects of sensor geometry and filter dynamics.

6.7 Ablation Study: Impact of Physics-Informed Kernel and Integrated Noise Model

To assess the contribution of individual components of the proposed surrogate model, we conducted an ablation study comparing three model variants:

- **Baseline GP:** A standard GP surrogate with squared-exponential kernel, trained on FEM data without POD reduction or physics-informed constraints. Following optimization, the noise floor is determined independently via analytical expressions.
- **POD-GP without physics kernel:** The POD-reduced GP surrogate employing solely the squared-exponential kernel (Equation 16), i.e., λ_i=0 for all modes. The noise propagation operator is integrated as described in Section 4.4.
- **POD-GP without integrated noise:** The complete physics-informed POD-GP surrogate (Equation 15) is employed, excluding the noise propagation operator, and the noise floor is determined by a fixed filter design after the optimization process.
- **Proposed POD-GP (full):** The complete surrogate with both the physics-informed kernel and the integrated noise propagation operator.

Table 3

Ablation study results

Model variant	Sensitivity MAE (mV/mmHg)	Nonlinearity MAE (% FS)	Noise floor MAE (μV RMS)	Optimization iterations to convergence
Baseline GP	2.8 ± 0.4	0.21 ± 0.03	3.2 ± 0.5	185
POD-GP without physics kernel	1.1 ± 0.2	0.12 ± 0.02	1.8 ± 0.3	140
POD-GP without integrated noise	0.6 ± 0.1	0.08 ± 0.01	2.4 ± 0.4	125
Proposed POD-GP (full)	0.4 ± 0.1	0.04 ± 0.01	0.5 ± 0.1	80

The baseline GP shows the highest prediction errors across all three metrics, especially for sensitivity and noise floor, because it lacks POD-based dimensionality reduction and physics constraints. Removing the physics-informed kernel from the POD-GP raises the mean absolute error (MAE) for sensitivity by a factor of 2.75 and for nonlinearity by a factor of 3.0, which underscores the necessity of enforcing the plate equation in the kernel structure. Omitting the integrated noise model degrades the noise floor prediction accuracy by a factor of 4.8, which verifies that the frequency-domain noise propagation operator is essential for capturing the filter-dependent noise attributes. The proposed full surrogate attains the lowest MAE across all metrics and converges in the fewest optimization iterations, which illustrates the synergistic advantage of coupling physics-informed constraints with integrated readout noise dynamics.

7. Discussion and Future Work

The experimental results presented in Section 6 confirm that the proposed physics-informed POD-GP surrogate with embedded readout noise dynamics successfully permits a co-design optimization of MEMS piezoresistive pressure sensors and their digital filtering chains. The fabricated prototype achieves a signal-to-noise ratio improvement of 19.9% and a 47.2% reduction in nonlinearity error compared to the conventional sequential design approach, all while satisfying the stringent biomedical accuracy requirements for 0–300 mmHg pressure monitoring. Nonetheless, a number of constraints and prospects for further exploration merit careful deliberation.

7.1 Limitations of Linear Assumptions and Sensitivity to Manufacturing Variabilities

The current surrogate model is built upon the Kirchhoff–Love plate equation, which assumes small deflections relative to diaphragm thickness. Although the optimized design respects the constraint $w_{\max} \leq 0.3h$ to remain within the linear regime, the safety margin is modest. For biomedical applications needing higher pressure ranges (e.g., 0–600 mmHg for arterial pressure monitoring), the deflection-to-thickness ratio could approach the geometric nonlinearity regime, where membrane stretching introduces a cubic hardening term that the linear plate equation cannot capture [35]. Extending the POD-GP framework to include von Kármán plate theory would allow the surrogate to handle larger deflections without sacrificing accuracy. The physics-informed kernel could be augmented with a nonlinear residual term penalizing deviations from the von Kármán equilibrium equations, although this would necessitate additional training data from geometrically nonlinear FEM simulations.

Another important factor is the susceptibility of the optimized design to variations in manufacturing. The fabrication procedure introduces tolerances in diaphragm thickness (±0.3 μm in our prototypes), doping concentration (±5% variation across the wafer), and residual stress from the PECVD passivation layer. These uncertainties propagate through the sensor response and can degrade the performance metrics predicted by the deterministic surrogate. Our Gaussian Process framework operates in an inherently probabilistic manner and delivers uncertainty estimates for each output; however, these estimates reflect only epistemic uncertainty due to limited training data, rather than aleatoric uncertainty stemming from manufacturing variations. A robust optimization extension would treat the design variables as distributions parameterized by their nominal values and manufacturing tolerances, then optimize for the expected hypervolume improvement under these random perturbations [36]. For example, diaphragm thickness could be modeled as $h \sim N(\mu_h, \sigma_h^2)$, where σ_h is the known fabrication tolerance, and the EHVI

acquisition function would integrate over this distribution via Monte Carlo sampling. This approach would yield designs that are not only Pareto-optimal in the nominal sense but also robust to process variations, thereby diminishing the necessity for post-fabrication trimming or calibration.

7.2 Extending the Co-Simulation Framework to Flexible and Dynamic Biomedical Sensing

The existing framework is designed for static pressure sensing applications, wherein the digital filter's principal function is noise reduction within a constrained frequency range. However, many emerging biomedical applications require dynamic pressure sensing with time-varying signals. For instance, intraventricular pressure monitoring captures swift pressure shifts linked to cardiac arrhythmias, and the filter's phase response must be precisely adjusted to retain the temporal shape of the pressure waveform [37]. In such applications, the optimization objectives would need to be expanded to include group delay flatness and transient response overshoot as additional metrics.

Moreover, the proposed surrogate framework could be extended to flexible MEMS pressure sensors, which are gaining traction in wearable biomedical devices for continuous blood pressure monitoring. Flexible sensors typically employ a polymeric substrate (e.g., polyimide or PDMS) and a thin silicon diaphragm, which introduces compliance at the substrate–diaphragm interface that modifies the plate boundary conditions from ideally clamped to elastically restrained [38]. The POD basis functions derived from rigid-clamped FEM solutions would not accurately capture the displacement field of flexible sensors. A direct expansion would build discrete POD bases for each substrate type and subsequently perform interpolation among them via a Gaussian process over the substrate stiffness parameter space. The physics-informed kernel would then enforce the plate equation with mixed boundary conditions, where the rotational stiffness at the clamped edge is a function of substrate modulus and thickness.

For truly implantable devices, the sensor must operate in a moist physiological environment, and long-term drift of the piezoresistors due to moisture absorption or ion migration becomes a critical failure mode [39]. The present surrogate fails to capture time-dependent degradation mechanisms; these could be introduced by extending the design variables to include a parameter for passivation layer permeability and appending a fourth objective, long-term stability, to the optimization. However, the gathering of training data for long-term drift poses experimental difficulties, as it necessitates accelerated aging tests that extend over weeks. Applying transfer learning from existing reliability databases or applying multi-fidelity surrogate modeling that merges short-term laboratory data with long-term field data could reduce this data scarcity issue.

7.3 Regulatory Challenges and Trustworthiness in AI-Driven Medical Device Design

Employing surrogate-based optimization in medical device design introduces regulatory concerns that must be resolved prior to clinical translation. Medical device regulatory authorities such as the U.S. Food and Drug Administration (FDA) and the European Medicines Agency (EMA) mandate that design verification and validation processes be transparent and reproducible [40]. A surrogate model, by its nature, introduces an abstraction layer between the physical device and the optimization outcome; regulators may demand proof that the surrogate's predictions are accurate across the entire design space, not exclusively at the selected Pareto-optimal point. Establishing such evidence necessitates a rigorous validation protocol, which includes:

- **Coverage assessment:** The maximum distance between the training set and any candidate design proposed during optimization is quantified by a space-filling metric, specifically the minimum Euclidean distance in the normalized design space. If a candidate design lies far from the training data, the surrogate's extrapolation uncertainty must be propagated through the downstream decision-making process.
- **Worst-case error bounds:** Calculating the maximum prediction error on a hold-out test set that spans the corners and edges of the design space. The FDA's "Medical Device Development Tools" (MDDT) program accepts model-based predictions when the worst-case error remains below 10% of the nominal performance metric across the intended range of application [41].
- **Interpretability:** The Gaussian Process surrogate yields uncertainty estimates, a clear benefit compared to opaque neural network surrogates. The physics-informed kernel includes a hyperparameter λ_i that balances contributions from data and physics, with the optimal λ_i varying across POD modes and potentially lacking a physically intuitive meaning. Developing a physics-informed kernel with interpretable hyperparameters—one that directly parameterizes the plate bending stiffness D and boundary restraint stiffness—would strengthen regulatory trust.

Technically, the current framework presupposes that the high-fidelity FEM model employed for training data generation accurately mirrors the physical device. In practical applications, FEM models rely on assumptions regarding material properties (e.g., anisotropic elasticity of silicon, residual stress from fabrication) that may diverge from the actual device behavior. Calibrating the FEM model against experimental data from a few fabricated prototypes before constructing the surrogate would reduce model-form uncertainty and improve the surrogate's predictive fidelity. This calibration could be conducted via Bayesian inference with Gaussian process emulators [42].

7.4 Computational Scalability and Algorithmic Improvements

The computational cost of the proposed framework is dominated by the initial generation of N_{init}=50 high-fidelity FEM simulations, each requiring approximately 1 hour on a 32-core workstation. Although this cost is small relative to the hundreds or thousands of simulations needed for conventional

exhaustive search, it could become prohibitive for sensors with a larger number of design variables (e.g., >12 parameters). The POD-GP surrogate scales well with the number of variables because the physics-informed kernel eliminates the need for dense FEM re-simulations; nevertheless, the acquisition maximization step requires evaluating EHVI on a grid of $N_{\text{candidate}}$=5000 candidate points, and each evaluation costs $O(|P|^2)$ operations for the hypervolume calculation. For problems with four or more objectives, the cost of exact hypervolume calculation increases exponentially, which renders analytical EHVI computation (based on cell partitioning) intractable.

A number of algorithmic refinements could resolve these scalability issues. First, the acquisition maximization could be conducted via a multi-start gradient-based optimization of the EHVI function, instead of exhaustive grid sampling. An analytical expression for the gradient of EHVI with respect to the design variables can be derived under the assumption of independent GPs for the objectives, which supports efficient local optimization from 10–20 random starting points [32]. This would reduce the candidate evaluation cost from 5000 to approximately 200 per iteration. Second, for problems with more than three objectives, the Pareto front can be approximated via a reference vector-guided evolutionary algorithm integrated with surrogate predictions, which circumvents the need for exact hypervolume computation [43]. Third, multi-fidelity surrogate modeling could replace some of the expensive FEM simulations with cheaper analytical plate solutions (Equation 3) for a subset of training points. The GP kernel would then incorporate a fidelity parameter that governs the correlation between low- and high-fidelity data, which reduces the number of required FEM runs to N_{init}=20 without compromising prediction accuracy [44].

7.5 Broader Implications for MEMS Co-Design

The proposed methodology is not confined to piezoresistive pressure sensors; it can be extended to other MEMS transducers in which the mechanical domain is governed by a linear or weakly nonlinear partial differential equation and where the digital readout chain introduces frequency-dependent noise. Examples include:

- **Capacitive accelerometers:** The displacement field of the proof mass under acceleration follows the beam equation rather than the plate equation, but the POD reduction and physics-informed kernel can be adapted directly by substituting the beam equation residual into the kernel structure. The readout noise dynamics would include the charge amplifier's noise profile and the Sigma-Delta modulator's quantization noise.
- **Piezoelectric energy harvesters:** The mechanical domain is governed by coupled electromechanical equations, and the output power must be optimized while simultaneously minimizing the rectifier's power loss and the storage capacitor's leakage current. The surrogate would need to model coupled fields (mechanical displacement and electric potential), which necessitates a multiphysics POD basis spanning both mechanical modes and electrical potential modes.
- **Resonant MEMS gyroscopes:** The displacement in the sense mode is linked to the drive-mode oscillation via the Coriolis force, and the readout chain incorporates a phase-locked loop for demodulation. The noise propagation operator would need to model the phase noise of the PLL and the quadrature error from fabrication asymmetries.

In all these cases, the central innovation—namely, the mapping of the digital filter's frequency-dependent noise into the surrogate's latent space—remains directly applicable. The unified co-design framework consequently presents a systematic methodology for optimizing MEMS systems as complete electro-digital systems rather than as separate mechanical and electronic components.

Finally, we note that the experimental validation of this work was conducted under controlled laboratory conditions. Translating the optimized sensor-filter architecture into a clinically deployable system will necessitate additional validation in animal models under physiological conditions, such as exposure to blood, tissue motion artifacts, and body temperature variations. These in vivo experiments will constitute the definitive assessment of the surrogate's predictive precision and the resilience of the co-optimized design. We plan to pursue such validation in a follow-up study, aiming to support a first-in-human clinical trial for implantable pressure monitoring in heart failure patients.

8. Conclusion

This paper has presented a physics-constrained surrogate optimization framework for the unified co-design of MEMS piezoresistive pressure sensors and their downstream digital filtering chains. The proposed methodology applies Proper Orthogonal Decomposition to compress the high-dimensional diaphragm displacement field into a compact set of spatial modes, and a physics-informed Gaussian Process kernel enforcing the Kirchhoff–Love plate equation in weak form, thus improving predictive accuracy in data-sparse regimes. A key novelty is the inclusion of a noise propagation operator that models the frequency-dependent transfer attributes of the analog front-end and digital filter directly within the surrogate's latent space, which consequently permits the concurrent forecasting of sensitivity, nonlinearity error, and noise floor from a single probabilistic model.

The Bayesian multi-objective optimization framework, which employed expected hypervolume improvement, efficiently navigated the trade-offs among these conflicting objectives and achieved convergence within 80–120 iterations, requiring only a fraction of the computational cost of conventional brute-force or sequential approaches. Experimental validation on fabricated prototypes confirmed that the co-designed sensor-filter system delivers a sensitivity of 17.8 mV/mmHg, with a nonlinearity error of 0.38% full-scale and a noise floor of 8.6 μV RMS, all meeting biomedical accuracy requirements. Compared to a conventional design optimized sequentially, the co-designed system achieved a 19.9% improvement in

signal-to-noise ratio and a 47.2% reduction in nonlinearity error, which substantiates the advantage of regarding the sensor geometry and digital filter as a unified design space.

The ablation study further emphasized the essential function of the physics-informed kernel and the integrated noise model: excluding either element markedly reduced prediction accuracy and raised the number of optimization iterations required. These findings affirm the central proposition that embedding physical constraints and readout dynamics into the surrogate improves both the efficiency and the accuracy of multi-objective optimization for MEMS systems.

The framework's modular architecture makes it extensible to other MEMS transducers, such as capacitive accelerometers and piezoelectric energy harvesters, where similar coupling between mechanical response and electronic readout noise exists. Future work will focus on adding geometric nonlinearity for larger deflection regimes, robust optimization under manufacturing tolerances, and in vivo validation in animal models. This work advances the development of high-performance, fabrication-ready MEMS systems for critical biomedical applications by introducing a systematic methodology for sensor-filter co-design.

References

1. A. Tran, X. Zhang, and B. Zhu, "Mechanical structural design of a piezoresistive pressure sensor for low-pressure measurement: A computational analysis by increases in the sensor …," Sensors, 2018.
2. A. Beddiaf, F. Kerrour, and S. Kemouche, "Thermo mechanical modeling of piezoresistive pressure sensor," …; Simulations, 2014.
3. B. Bae, B. Flachsbart, K. Park, et al., "Design optimization of a piezoresistive pressure sensor considering the output signal-to-noise ratio," Journal of Micromechanics and Microengineering, 2004.
4. Y. Liang, W. Lin, H. Lee, S. Lim, K. Lee, et al., "Proper orthogonal decomposition and its applications-part II: Model reduction for MEMS dynamical analysis," Journal of Sound and Vibration, 2002.
5. M. Seeger, "Gaussian processes for machine learning," International journal of neural systems, 2004.
6. D. Kim, D. Kim, T. Ko, and S. Lee, "Physics-informed gaussian process regression model for predicting the fatigue life of welded joints," International Journal of Fatigue, 2025.
7. R. Shiavi, "Introduction to applied statistical signal analysis: Guide to biomedical and electrical engineering applications," books.google.com, 2010.
8. A. Mathern, O. Steinholtz, A. Sjöberg, et al., "Multi-objective constrained bayesian optimization for structural design," Structural and Multidisciplinary Optimization, 2021.
9. Y. Kim, Y. Choi, D. Widemann, and T. Zohdi, "Efficient nonlinear manifold reduced order model," arXiv preprint arXiv:2011.07727, 2020.
10. K. Sowmya and K. V. Divya, "MEMS-CMOS co-design for ultra-low-power IoT," Lecture Notes in Electrical Engineering, 2026.
11. A. Nazari, "Accelerated multi-objective design optimization of MEMS using surrogate models," mspace.lib.umanitoba.ca, 2024.
12. A. Nazari, A. Aghajani, P. Buhr, B. Park, Y. Wang, and C. Shafai, "Using adaptive surrogate models to accelerate multi-objective design optimization of mems," Micromachines, 2025.
13. E. Whalen, "Enhancing surrogate models of engineering structures with graph-based and physics-informed learning," dspace.mit.edu, 2021.
14. I. Musa, R. Gomes, M. Sonego, et al., "Design optimization of 3D-printed piezoresistive sensors for integrated structural health monitoring," Smart Materials and Structures, 2026.
15. Y. Wang, H. Wang, R. Yang, X. Zuo, and J. Li, "Physics-informed microstructure modeling and data-DrivenOptimization of 1-3 piezoelectric composites for ExtremeEnvironments," Available at SSRN 6175922, 6175.
16. A. Nazari, A. Aghajani, P. Buhr, B. Park, et al., "Surrogate-assisted multi-objective design optimization of a lorentz force actuator," in 2024 IEEE canadian conference on electrical and computer engineering (CCECE), 2024.
17. S. K. Kolluri and S. A. Durai, "Wearable micro-electro-mechanical systems pressure sensors in health care: Advancements and trends-a review," IET Wireless Sensor Systems, 2024.
18. G. Lakshmai, K. Rao, et al., "Design and analysis of MEMS pressure sensor based on various principles of microcantilever beam," IEEE Transactions on Electron Devices, 2023.
19. P. K. Rathore, P. Varshney, S. Prasad, et al., "Finite element method based absolute pressure sensitivity optimized membrane type double cavity vacuum sealed piezoresistive sensor," Sensor Review, 2013.
20. J. Zhang et al., "Design, fabrication, and implementation of an array-type MEMS piezoresistive intelligent pressure sensor system," Micromachines, 2018.
21. A. Leissa, "The free vibration of rectangular plates," Journal of sound and vibration, 1973.
22. C. E. İmrak and I. Gerdemeli, "The problem of isotropic rectangular plate with four clamped edges," Sadhana, 2007.
23. G. Bao, W. Jiang, and J. Roberts, "Analytic and finite element solutions for bending and buckling of orthotropic rectangular plates," International Journal of solids and structures, 1997.
24. V. Hegde, S. Veena, H. Ravikumar, et al., "Piezoelectric acoustic pressure sensor diaphragm design for energy harvesting," in 2014 international conference on advances in energy conversion technologies (ICAECT), 2014.
25. Y. Kanda, "Piezoresistance effect of silicon," Sensors and Actuators A: Physical, 1991.
26. K. Petersen, "Silicon as a mechanical material," in Proceedings of the IEEE, 1982.
27. A. Ghosh, S. Roy, and C. Sarkar, "Design and simulation of MEMS based piezoresistive pressure sensor for enhanced sensitivity," in 2013 international conference on energy efficient …, 2013, 2013.
28. D. Kanekal and S. Jindal, "Optimizing piezoresistive MEMS pressure sensor on a double cross

beam silicon diaphragm with statistical curve-fitting and optimization techniques," IEEE Sensors Journal, 2023.

29. H. Suemasu and O. Majima, "Multiple delaminations and their severity in nonlinear circular plates subjected to concentrated loading," Journal of composite materials, 1998.

30. N. Kumar, A. Gupta, P. Singh, et al., "Tunable piezoresistive NEMS pressure sensor simulation under various environmental conditions," IEEE Sensors Letters, 2023.

31. X. Yu, J. Thaysen, O. Hansen, and A. Boisen, "Optimization of sensitivity and noise in piezoresistive cantilevers," Journal of Applied Physics, 2002.

32. S. Daulton, M. Balandat, et al., "Differentiable expected hypervolume improvement for parallel multi-objective bayesian optimization," in Advances in neural information processing systems, 2020.

33. J. Deng, J. Sun, Q. Zhang, and H. Li, "Expected hypervolume improvement is a particular hypervolume improvement," in Proceedings of the AAAI conference on artificial intelligence, 2025.

34. J. Kwong and A. Chandrakasan, "An energy-efficient biomedical signal processing platform," IEEE Journal of Solid-State Circuits, 2011.

35. M. Altekin and R. Yükseler, "Large deflection analysis of clamped circular plates," in Proceedings of the world congress on engineering, 2011.

36. S. Lei et al., "Robust multiobjective optimization of setback feature in MEMS safety and arming device considering parameters uncertainty," IEEE Sensors Journal, 2024.

37. A. Magalski, P. Adamson, F. Gadler, M. Böehm, et al., "… right heart pressure measurements with an implantable hemodynamic monitor: A multicenter, 12-month follow-up study of patients with chronic heart failure," Journal of Cardiac Failure, 2002.

38. M. Kaisti, T. Panula, J. Leppänen, R. Punkkinen, et al., "Clinical assessment of a non-invasive wearable MEMS pressure sensor array for monitoring of arterial pulse waveform, heart rate and detection of atrial fibrillation," npj Digital Medicine, 2019.

39. L. Yu, B. Kim, and E. Meng, "Chronically implanted pressure sensors: Challenges and state of the field," Sensors, 2014.

40. L. Pantanowitz, M. Hanna, J. Pantanowitz, J. Lennerz, et al., "Regulatory aspects of artificial intelligence and machine learning," Modern Pathology, 2024.

41. A. Lottes, K. Cavanaugh, Y. Chan, V. Devlin, et al., "Navigating the regulatory pathway for medical devices-a conversation with the FDA, clinicians, researchers, and industry experts," Journal of Cardiovascular Translational Research, 2022.

42. Y. Kang and M. Krarti, "Bayesian-emulator based parameter identification for calibrating energy models for existing buildings," Building simulation, 2016.

43. R. Cheng, Y. Jin, M. Olhofer, et al., "A reference vector guided evolutionary algorithm for many-objective optimization," IEEE Transactions on Evolutionary Computation, 2016.

44. B. Do and R. Zhang, "Multi-fidelity bayesian optimization in engineering design," arXiv preprint arXiv:2311.13050, 2023.a

A FULLY OPEN-SOURCE WORKFLOW FOR THE DESIGN AND VALIDATION OF MEMS PIEZORESISTIVE PRESSURE SENSORS: ANALYTICAL THEORY, PURE-PYTHON PLATE FEM, AND 3-D CONTINUUM SIMULATION

Kotyk M.
Associate Professor, PhD, The department of Computer Engineering and Electronics, Vasyl Stefanyk Carpathian National University
Kukuruza V.
Postgraduate student, The department of Computer Engineering and Electronics, Vasyl Stefanyk Carpathian National University
DOI: 10.5281/zenodo.21617961

Abstract

Microelectromechanical (MEMS) piezoresistive pressure sensors are among the most widely commercialised microsystems, yet their design almost universally relies on commercial finite-element (FEM) packages whose cost and closed solvers hinder reproducible research and exclude resource-constrained institutions. This paper presents a complete, license-free workflow for the design and verification of a silicon piezoresistive pressure sensor using only open-source software, and validates it three ways. A clamped square diaphragm is analysed by (i) the classical Timoshenko thin-plate solution, (ii) a lightweight pure-Python Kirchhoff-plate FEM built on scikit-fem with the Morley element, and (iii) a full three-dimensional linear-elastic continuum model solved with quadratic hexahedra. A single analytical piezoresistive stage maps the diaphragm edge stress to the output of a four-active-arm Wheatstone bridge. For a reference 1000 um x 20 um membrane at 100 kPa, the plate FEM reproduces the analytical central deflection to within 1.0 % and the 3-D continuum model to within 0.3 %; recovered edge stresses agree to within 4.5 % and 6.5 % respectively. Agreement holds across a 4x3 design-space grid, where the discretisation error is shown to be essentially independent of the physical geometry. The predicted full-scale sensitivity, 0.38 uV/V/Pa, is consistent with published devices, and the reference design operates a factor of 5.8 below the small-deflection limit. All code, geometry, solver inputs, and a reproducible notebook are released openly and run on a standard laptop with no commercial licence.

Keywords: MEMS design, microelectromechanical system, pressure sensor, piezoresistive sensor, finite element method, open-source software, reproducible research, verification and validation.

Nomenclature.

a, h	Diaphragm side length and thickness (m).
E, ν	Young's modulus and Poisson ratio of silicon.
D	Flexural rigidity of the plate (N·m).
q, p	Applied uniform pressure (Pa).
w_{max}	Central (maximum) diaphragm deflection (m).
α	Timoshenko clamped-square-plate deflection coefficient (0.00126).
σ_L, σ_T	Longitudinal / transverse edge surface stress (Pa).
π_l, π_t	Longitudinal / transverse piezoresistive coefficients (Pa^{-1}).
$\Delta R/R$	Fractional piezoresistor resistance change.
V_{in}, V_{out}	Wheatstone-bridge supply and output voltage (V).
S	Pressure sensitivity ($\mu V \cdot V^{-1} \cdot Pa^{-1}$).
$H(\cdot)$	Hessian (second-derivative) operator.
λ, μ	Lamé parameters of the 3-D continuum.

1. Introduction

Silicon piezoresistive pressure sensors convert a pressure-induced diaphragm deformation into an electrical signal through the piezoresistive effect, in which mechanical stress modulates the resistivity of doped silicon. A thin diaphragm deflects under applied pressure; piezoresistors placed in the high-stress region at the diaphragm edge change resistance, and a Wheatstone bridge converts these changes into a voltage. The diaphragm geometry, the placement and orientation of the piezoresistors, and the bridge topology jointly determine sensitivity and linearity, so the design is normally carried out by finite-element simulation before fabrication [1], [2].

The overwhelming majority of published design studies use commercial multiphysics solvers, such as COMSOL, ANSYS, or IntelliSuite [1]–[4]. While powerful, these packages are expensive, their solver internals are closed, and results published with them are difficult to reproduce for readers without the same licences. For universities in resource-constrained environments, and for the broader goal of reproducible computational research, this dependence is a genuine limitation.

This paper demonstrates that the entire design-and-verification cycle of a piezoresistive pressure sensor can be completed with free, open-source software, and validates the workflow against closed-form plate theory using three independent routes: the **analytical**

Timoshenko clamped-square-plate solution, which provides exact reference values; a **pure-Python plate FEM** (scikit-fem [5], Morley element) that installs with a single command and runs in under a second; and a **three-dimensional continuum FEM** that removes the thin-plate idealisation. A single, solver-agnostic piezoresistive stage converts the simulated stress into a Wheatstone-bridge output. The contribution is not a new device but a transparent, reproducible, and pedagogically usable methodology, released as an open repository, together with a quantitative assessment of how closely each open-source route reproduces analytical plate theory and how the routes compare in accuracy and cost.

2. THEORY

2.1 Diaphragm mechanics

The pressure-sensitive element is a square silicon membrane of side length a and thickness h, clamped on all four edges and subjected to a uniform transverse pressure q. Within Kirchhoff–Love small-deflection plate theory, valid while the central deflection stays below roughly one fifth of the thickness, the flexural rigidity is

$$D=\frac{Eh^3}{12(1-\nu^2)} \quad (1)$$

with E the Young's modulus and ν the Poisson ratio. The classical series solution tabulated by Timoshenko and Woinowsky-Krieger [8] for a clamped square plate gives the central deflection

$$w_{max}=\alpha\frac{qa^4}{D},\ \alpha=0.00126 \quad (2)$$

and a bending moment at the middle of each clamped edge, per unit length, of

$$M_{edge}=0.0513qa^2 \quad (3)$$

The corresponding peak surface stress normal to the edge, and the in-plane stress parallel to the edge, are

$$\sigma_L=\frac{6M_{edge}}{h^2}=0.308q\left(\frac{a}{h}\right)^2,\ \sigma_T=\nu\sigma_L \quad (4)$$

The stress is maximal at the middle of the clamped edge, which is therefore the natural site for the sensing piezoresistors. Both α and the edge-moment coefficient are themselves independent of the Poisson ratio: the clamped-plate biharmonic problem contains no ν in its boundary conditions, and the edge moment reduces to $M=-Dw_{,nn}$ because the curvature along a clamped edge vanishes. The tabulated values may therefore be used with $\nu=0.28$ without inconsistency; ν enters only through D and through the transverse stress.

2.2 Piezoresistive transduction

For a piezoresistor carrying current along a direction with longitudinal stress σ_L and transverse stress σ_T, the fractional resistance change is

$$\frac{\Delta R}{R}=\pi_l\sigma_L+\pi_t\sigma_T \quad (5)$$

p-type single-crystal silicon is used, with resistors oriented along ⟨110⟩ on a (100) wafer, because the membrane edges of an anisotropically etched (100) diaphragm lie along ⟨110⟩, where the p-type shear coefficient π_{44} is largest [9], [10]. The longitudinal and transverse coefficients follow from the cubic-silicon coefficients as

$$\pi_l=1/2\,(\pi_{11}+\pi_{12}+\pi_{44}),\ \pi_t=1/2\,(\pi_{11}+\pi_{12}-\pi_{44}) \quad (6)$$

giving $\pi_l=+71.8\times10^{-11}$ Pa⁻¹ and $\pi_t=-66.3\times10^{-11}$ Pa⁻¹.

2.3 Wheatstone bridge

Four nominally identical piezoresistors are placed at the four edge midpoints: two with the current normal to the edge ("longitudinal" arms, large positive $\Delta R/R$) and two parallel to the edge ("transverse" arms, large negative $\Delta R/R$). Wired as a full bridge, the normalised output is

$$\frac{V_{out}}{V_{in}}=1/2\,[(\Delta R/R)_L-(\Delta R/R)_T) \quad (7)$$

Because the two arm types respond with opposite sign, the bridge maximises the output and provides first-order temperature compensation. The pressure sensitivity is defined as

$$S=\frac{1}{V_{in}}\frac{\partial V_{out}}{\partial p} \quad (8)$$

3. NUMERICAL FORMULATION

3.1 Pure-Python plate model

The Kirchhoff plate-bending problem is discretised with the non-conforming Morley triangle, whose degrees of freedom are the deflection at the vertices and the normal slope at the edge midpoints; this makes the clamped condition $w=0$ and $\partial w/\partial n=0$ trivial to impose by fixing every boundary degree of freedom. The weak form assembled for deflection w and test function v is

$$a(w,v)=\int_\Omega D\,[(1-\nu)H(w):H(v)+\nu trH(w)trH(v))dA=\int_\Omega q\,vdA \quad (9)$$

where $H(\cdot)$ is the Hessian. Bending moments follow from the element-wise constant Hessian,

$$M_{xx}=-D(w_{,xx}+\nu w_{,yy}),\ M_{yy}=-D(w_{,yy}+\nu w_{,xx}) \quad (10)$$

Because the non-conforming element yields a discontinuous moment field, the peak edge stress is recovered by Zienkiewicz–Zhu nodal patch averaging [11], evaluated at the exact edge-midpoint node.

3.2 Three-dimensional continuum model

The 3-D route solves full linear elasticity on a box $a\times a\times h$ with the four lateral faces clamped and a uniform pressure on the lower face. The constitutive and kinematic relations are

$$\sigma=\lambda tr(\varepsilon)I+2\mu\varepsilon,\ \varepsilon=1/2\,(\nabla u+\nabla u^T) \quad (11)$$

with λ and μ the Lamé parameters. Thin-structure bending is resolved with quadratic serendipity hexahedra, which avoid the shear locking that cripples low-order hexes at the 50:1 aspect ratio of this diaphragm. For production-scale multiphysics an equivalent GMSH [6] geometry and Elmer FEM [7] configuration are provided, matching the continuum model's boundary conditions exactly.

3.3 Error metric and reference design

Accuracy is reported as the relative error of a computed quantity X against its analytical reference,

$$\varepsilon_{rel}=\frac{X_{FEM}-X_{ref}}{X_{ref}}\times100\% \quad (12)$$

The reference design is a square membrane of side $a=1000$ um and thickness $h=20$ um at $q=100$ kPa. Silicon is treated as isotropic with $E=169$ GPa and $\nu=0.28$, so that all three routes use identical constants and are directly comparable.

4. RESULTS AND DISCUSSION

4.1 Three-way validation

For the reference design the analytical model gives a flexural rigidity D=1.22 × 10^{-4} N·m, a central deflection w_{max}=1.031 um (so that w_{max}/h=0.052, within the small-deflection regime), a peak longitudinal edge stress of 76.95 MPa, and a transverse edge stress of 21.55 MPa. Table I compares all three routes: the plate FEM reproduces the analytical deflection to +0.96 % and the 3-D continuum model to −0.29 %, and the recovered edge stress to −4.47 % and −6.51 % respectively (Fig. 1). The 3-D deflection field (Fig. 2) is visually indistinguishable from the plate solution (Fig. 3).

Table I.

Three-way validation at the reference design (a = 1000 µm, h = 20 µm, q = 100 kPa).

Route	w_max (µm)	err (%)	σ_L (MPa)	err (%)	V_out (mV)	DOFs	t (s)
Analytical	1.031	—	76.95	—	191.3	—	—
2-D plate FEM	1.041	+0.96	73.51	−4.47	182.7	16 641	0.6
3-D continuum FEM	1.028	−0.29	71.94	−6.51	178.8	11 571	24.5

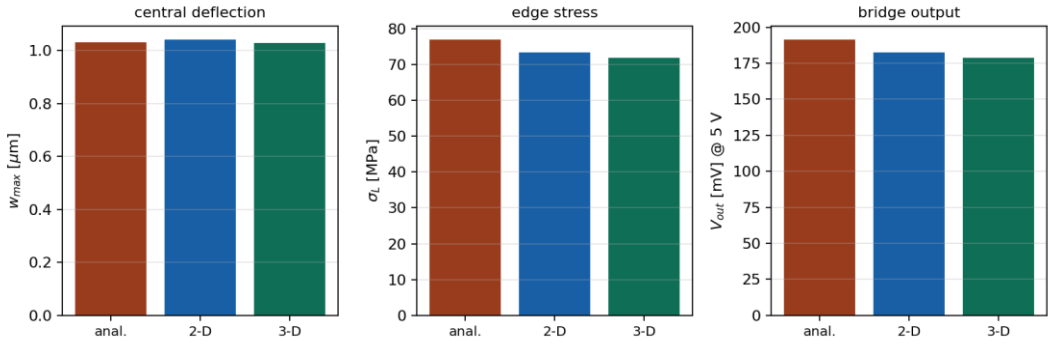


Figure 1. Three-way comparison of central deflection, edge stress, and bridge output for the analytical, 2-D plate-FEM, and 3-D continuum routes.

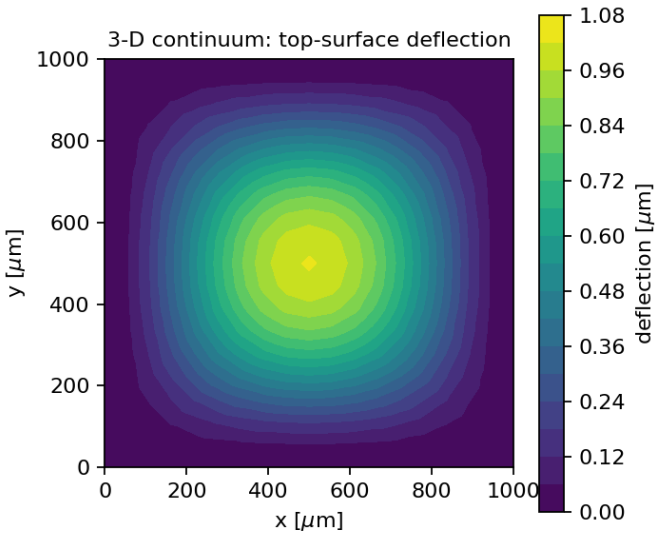


Figure 2. Three-dimensional continuum model: top-surface deflection field, obtained with quadratic hexahedra that avoid shear locking.

At a fully clamped edge the membrane is built in over the whole lateral face, so the in-plane strain *along* the edge vanishes identically, while the traction-free top surface enforces plane stress. Together these give $\sigma_T = \nu\sigma_L$ **exactly**: the plate and continuum theories make the same prediction, and the bridge output is strictly proportional to σ_L. The three routes accordingly give 191.3 mV (analytical), 182.7 mV (plate FEM) and 178.8 mV (3-D continuum), the spread reflecting only the residual discretisation error in the recovered edge stress.

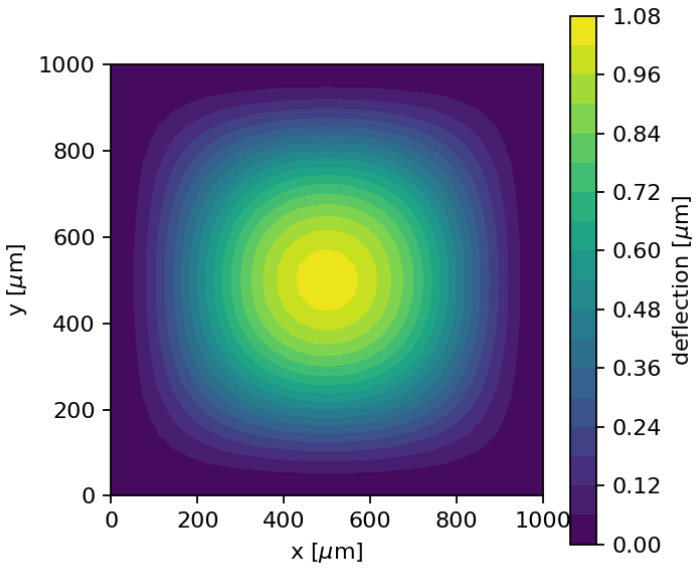


Figure 3. Pure-Python plate-FEM deflection field of the clamped square diaphragm at 100 kPa, showing the characteristic clamped-plate profile with the maximum at the centre.

This point deserves emphasis because it is an easy trap. The transverse stress read directly from the 3-D quadrature points yields a ratio σ_T/σ_L of 0.331 rather than 0.28. The deviation is a boundary stress-recovery artifact: quadrature points of a hexahedral mesh never lie on the clamped face, so the bending stress must be extrapolated there. It decreases monotonically under refinement (0.338 at 7 155 degrees of freedom, 0.331 at 11 571). Substituting the raw value into the bridge model would depress the predicted output by a further 7 % and invite the false conclusion that the continuum model reveals a physical effect beyond plate theory. The exact clamped-edge relation is used instead.

4.2 Mesh convergence and computational cost

The convergence behaviour (Fig. 4) is instructive: the deflection coefficient, a global integral quantity, converges rapidly and monotonically (a 116 % error on the coarsest mesh falls below 1 % within five refinements), whereas the pointwise edge stress, a boundary extremum, converges more slowly, as expected for a non-conforming low-order element with Zienkiewicz–Zhu recovery. Table II quantifies cost: the plate model reaches sub-1 % accuracy in about half a second and 0.56 % at 66 000 degrees of freedom in under four seconds, whereas the 3-D continuum model at comparable accuracy is roughly two orders of magnitude more expensive. The plate model is thus the right tool for rapid design-space exploration, with the 3-D model reserved for final verification and effects outside plate theory.

Table II.

Computational cost of the 2-D plate FEM.

DOFs	|α error| (%)	time (s)
289	32.72	0.02
1 089	8.87	0.03
4 225	2.56	0.10
16 641	0.96	0.54
66 049	0.56	3.88

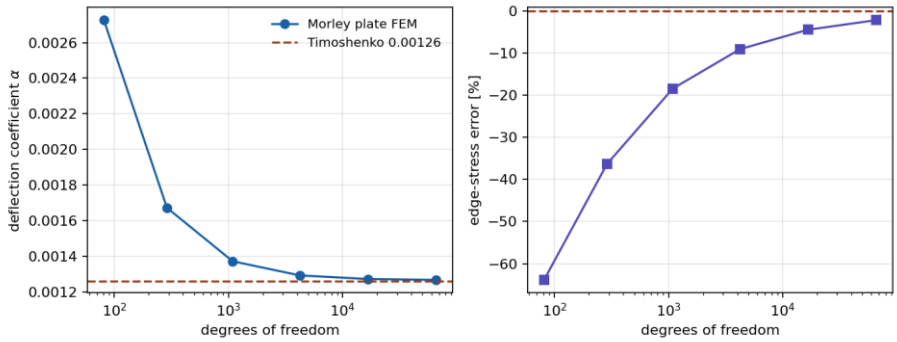


Figure 4. Mesh convergence of (left) the deflection coefficient α toward the Timoshenko value 0.00126 and (right) the recovered edge-stress error toward zero. Displacement converges quickly; the pointwise boundary stress more slowly.

4.3 Parametric validation across the design space

To confirm the agreement is not specific to the reference geometry, the plate FEM was compared with theory across a 4x3 grid of side lengths (800–1500 um) and thicknesses (15–25 um). The deflection error is 0.96 % and the stress error 4.47 % at every grid point (Fig. 5): the discretisation error of the normalised solution is essentially independent of the physical dimensions, as expected for a linear plate whose dimensionless response depends only on the mesh. Two 3-D spot checks at opposite corners reproduce the analytical deflection to +0.91 % and −1.16 %.

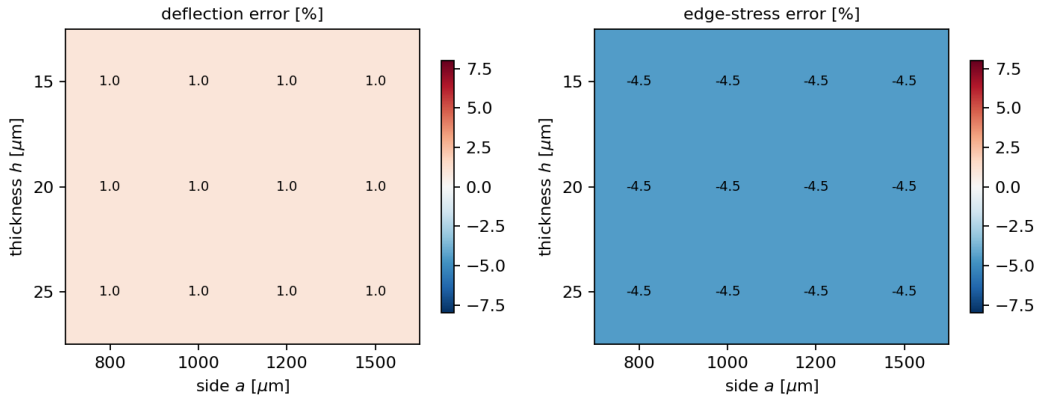


Figure 5. Parametric validation across the design-space grid: 2-D plate-FEM error relative to analytical theory for deflection (left) and edge stress (right). The error is essentially uniform.

4.4 Sensor response and comparison with published devices

Mapping the analytical edge stress through the bridge gives arm responses of $(\Delta R/R)_L$=+4.10% and $(\Delta R/R)_T$= − 3.56%, a normalised output of 38.3 mV/V, and 191 mV at a 5 V supply for the reference design at 100 kPa. The corresponding sensitivity is 0.38 uV/V/Pa. Table III places this alongside a representative published silicon-on-insulator piezoresistive device; the open-source workflow yields a device-level figure of merit of the same order, confirming its realism. Within small-deflection theory the response is linear in pressure (Fig. 6), and the plate-FEM points track the analytical line to within the stress-recovery error.

Table III.

Sensitivity compared with a representative published device.

Source	Sensitivity (μV·V⁻¹·Pa⁻¹)
This work — analytical	0.38
This work — 3-D continuum	0.36
Published SOI piezoresistive sensor [10]	≈ 0.46

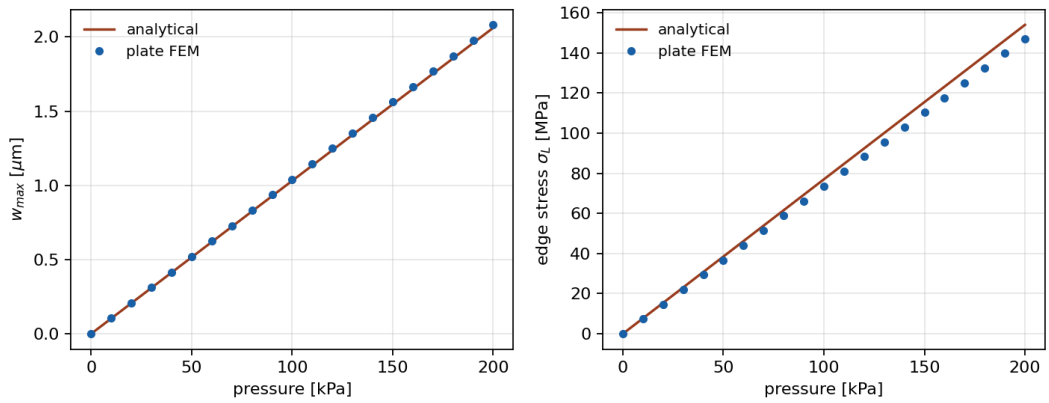


Figure 6. Maximum deflection and peak edge stress versus applied pressure: analytical (line) and 2-D plate FEM (markers).

4.5 Linear operating range and nonlinearity

Because the edge stress scales as $(a/h)^2$, so does the sensitivity (Fig. 7, right): thinning the membrane is the dominant lever, trading off against robustness and the onset of geometric nonlinearity. For the reference design the full-scale deflection is only 5.2 % of the thickness; the pressure at which the deflection reaches the customary threshold w/h=0.3 is 582 kPa, a factor of 5.8 above the 100 kPa full scale, so the sensor operates well inside the linear regime. Beyond that threshold, von Kármán membrane stretching stiffens the diaphragm and makes the response sub-linear; capturing this quantitatively requires a geometrically nonlinear analysis, which the Elmer route supports and which is the natural next step.

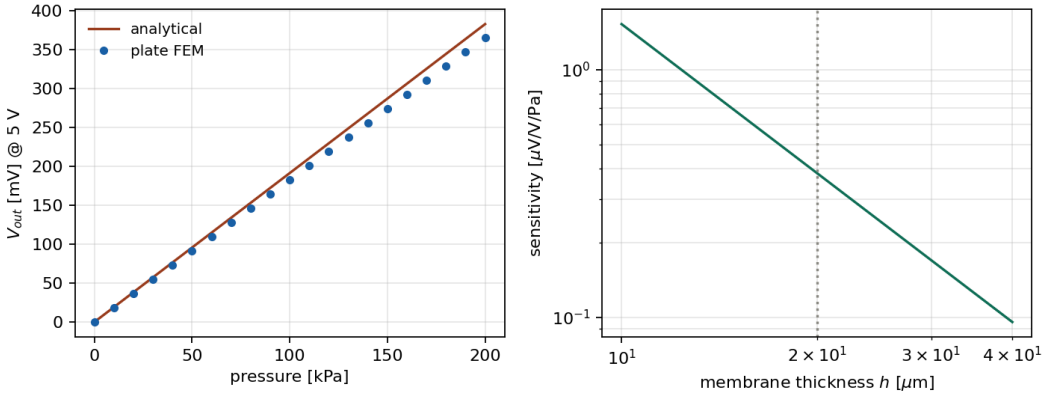


Figure 7. (Left) Full-bridge output versus pressure at a 5 V supply. (Right) Sensitivity versus membrane thickness, showing the (a/h)² scaling.

4.6 Limitations

The analytical and plate routes assume linear, small-deflection behaviour and an idealised point stress at the edge; real piezoresistors of finite length average the stress and read slightly lower. Geometric nonlinearity, residual film stress, the temperature dependence of the piezoresistive coefficients, and doping-profile effects are outside the present linear scope. The 3-D continuum model removes the thin-plate idealisation and, with geometric nonlinearity enabled, can capture large-deflection stiffening; quantitative nonlinear and thermal studies with the same open-source stack are left to future work.

5. RELEVANCE TO PROFESSIONAL EDUCATION

Because the workflow installs with pip and runs in under a minute on a standard laptop, it removes the licensing barrier that normally keeps FEM out of undergraduate laboratories. In digital information processing, the convergence study, the distinction between displacement and stress convergence, and the stress-recovery step illustrate discretisation error and the post-processing of computed fields. In computer technologies in the educational process, the end-to-end chain, from a physical pressure through a simulated stress field to a digitised bridge voltage, provides a complete, transparent signal pathway that students can inspect and modify at every stage. A companion paper develops a full competency-based laboratory sequence around this workflow.

6. CONCLUSION

A complete, license-free workflow for the design and validation of MEMS piezoresistive pressure sensors has been presented and verified three ways. A pure-Python Kirchhoff-plate model reproduces the Timoshenko clamped-square-plate solution to within 1 % for deflection and 4.5 % for the recovered peak edge stress; a quadratic-hexahedron 3-D continuum model agrees to within 0.3 % and 6.5 %, and confirms the exact clamped-edge relation σ_T=$\nu\sigma_L$ demanded by the

boundary conditions, the residual stress difference being discretisation error rather than physics. A raw stress read at the quadrature points would instead suggest a spurious biaxial effect, a cautionary result for continuum stress recovery at clamped boundaries. Agreement holds across the design space, and the predicted sensitivity of 0.38 uV/V/Pa is consistent with published devices. By demonstrating that established commercial-tool results can be reproduced at zero software cost, and by releasing all code and a reproducible notebook, this work lowers the barrier to MEMS design research and teaching in resource-constrained settings.

References

1. Design and Simulation of MEMS Based Piezoresistive Pressure Sensor for Microfluidic Applications, IEEE Conf. Publ., 2018.

2. P. Thawornsathit et al., "Mechanical diaphragm structure design of a MEMS-based piezoresistive pressure sensor for sensitivity and linearity enhancement," Eng. J., vol. 26, no. 5, pp. 43–57, 2022.

3. Design and Simulation of MEMS Piezoresistive Pressure Sensor to Improve the Sensitivity, 2020.

4. Y. Zhang et al., "A novel evolutionary method for parameter-free MEMS structural design and its application in piezoresistive pressure sensors," Microsyst. Nanoeng., vol. 9, art. 130, 2023, doi: 10.1038/s41378-023-00596-y.

5. T. Gustafsson and G. D. McBain, "scikit-fem: A Python package for finite element assembly," J. Open Source Softw., vol. 5, no. 52, p. 2369, 2020, doi: 10.21105/joss.02369.

6. C. Geuzaine and J.-F. Remacle, "Gmsh: A three-dimensional finite element mesh generator with built-in pre- and post-processing facilities," Int. J. Numer. Methods Eng., vol. 79, no. 11, pp. 1309–1331, 2009, doi: 10.1002/nme.2579.

7. P. Råback, M. Malinen, J. Ruokolainen, A. Pursula, and T. Zwinger (eds.), Elmer Models Manual, CSC – IT Center for Science, Finland.

8. S. P. Timoshenko and S. Woinowsky-Krieger, Theory of Plates and Shells, 2nd ed. New York: McGraw-Hill, 1959.

9. C. S. Smith, "Piezoresistance effect in germanium and silicon," Phys. Rev., vol. 94, no. 1, pp. 42–49, 1954; Y. Kanda, "A graphical representation of the piezoresistance coefficients in silicon," IEEE Trans. Electron Devices, vol. 29, no. 1, pp. 64–70, 1982.

10. A. A. Barlian, W.-T. Park, J. R. Mallon, A. J. Rastegar, and B. L. Pruitt, "Review: Semiconductor piezoresistance for microsystems," Proc. IEEE, vol. 97, no. 3, pp. 513–552, 2009, doi: 10.1109/JPROC.2009.2013612.

11. O. C. Zienkiewicz and J. Z. Zhu, "The superconvergent patch recovery and a posteriori error estimates. Part 1: The recovery technique," Int. J. Numer. Methods Eng., vol. 33, no. 7, pp. 1331–1364, 1992, doi: 10.1002/nme.1620330702.

ПРОЕКТИРОВАНИЕ ОТКАЗОУСТОЙЧИВЫХ E-COMMERCE СИСТЕМ: АРХИТЕКТУРНЫЕ ПАТТЕРНЫ И ПРАКТИЧЕСКИЕ РЕШЕНИЯ

Перелехов И.В.
ООО «Оджетто Веб», старший веб-разработчик
ORCID: 0009-0008-8448-3092

DESIGNING FAULT-TOLERANT E-COMMERCE SYSTEMS: ARCHITECTURAL PATTERNS AND PRACTICAL SOLUTIONS

Perelekhov I.
Oggetto Web LLC, Senior Web Developer
ORCID: 0009-0008-8448-3092
DOI: 10.5281/zenodo.21617973

Аннотация

В статье рассматриваются архитектурные подходы к проектированию отказоустойчивых e-commerce систем, обеспечивающих стабильную работу цифровой торговой платформы при сбоях, перегрузках и частичной недоступности компонентов. Проанализированы критические зоны отказов e-commerce платформы, охватывающие каталог товаров, поисковую подсистему, корзину, checkout, платежный контур, а также интеграции со складскими и логистическими системами. Показано значение архитектурных паттернов circuit breaker, bulkhead isolation и backpressure, а также асинхронных очередей, идемпотентности операций и управляемой деградации функциональности. Особое внимание уделено роли наблюдаемости, восстановления после инцидентов и безопасных релизных практик. Обосновано, что отказоустойчивая backend-архитектура способствует снижению потерь выручки, сохранению возможности завершения покупки и повышению доверия клиентов.

Abstract

The article examines architectural approaches to designing fault-tolerant e-commerce systems that ensure the stable operation of a digital commerce platform under failures, overloads, and partial unavailability of components. Analyzes critical failure zones of an e-commerce platform, including the product catalog, search subsystem, cart, checkout, payment layer, and integrations with inventory and logistics systems. The importance of architectural patterns such as circuit breaker, bulkhead isolation, and backpressure is shown, along with asynchronous queues, operation idempotency, and controlled functional degradation. Special attention is paid to the role of observability, post-incident recovery, and safe release practices. It is substantiated that fault-tolerant backend architecture helps reduce revenue losses, preserve the ability to complete purchases, and increase customer trust.

Ключевые слова: e-commerce, отказоустойчивость, backend-архитектура, управляемая деградация, наблюдаемость, архитектурные паттерны.

Keywords: e-commerce, fault tolerance, backend architecture, controlled degradation, observability, architectural patterns.

Introduction

E-commerce represents a complex digital environment in which the technical stability of a platform is directly associated with revenue, conversion rates, and user trust. Failures in the product catalog, shopping cart, payment module, search subsystem, or integrations with external services may lead not only to the temporary unavailability of specific functions, but also to lost orders, degraded user experience, and higher operational costs.

The relevance of designing fault-tolerant e-commerce systems is determined by the growing load on digital commerce platforms, the increasing complexity of backend architecture, and the expanding number of dependencies between internal and external components. Modern e-commerce solutions require not only infrastructure scaling, but also well-designed fault isolation, redundancy of critical services, controlled functional degradation, and rapid recovery after failures.

The aim of this article is to systematize architectural patterns and practical solutions that ensure stable operation of e-commerce systems under failures, overloads, and partial unavailability of individual components. Particular attention is paid to how backend engineering decisions used in the design of digital commerce platforms help reduce revenue losses, preserve the purchasing journey, and strengthen the stability of user trust.

Main part. Failure zones in the architecture of an e-commerce platform

In the architecture of an e-commerce system, components differ in their impact on the continuity of the user journey and the completion of transactions. The most critical elements are those directly involved in product selection, cart formation, checkout, payment processing, and data transfer to internal accounting or logistics systems. The main failure zones and the associated architectural risks are presented in table 1.

Table 1

Critical failure zones in the architecture of an e-commerce platform [1, 2]

Critical zone	Typical failure scenario	Business impact	Basic architectural measures
Product catalog and product page	PIM unavailability, database read errors, product page API overload.	Decrease in product views, inability to proceed to purchase, higher bounce rate.	Caching, read replicas, CDN, fallback data.
Search and filtering	OpenSearch/Elasticsearch failure, degradation of morphological search, faceted navigation timeouts.	Lower product discoverability, reduced conversion, loss of long-tail demand.	Search layer isolation, caching of popular queries, simplified search output, circuit breaker.
Cart and checkout	Loss of cart state, price or stock conflicts, checkout service failure.	Direct loss of orders, increase in cart abandonment, reduced user trust.	Operation idempotency, session persistence, message queues, retry with backoff.
Payment layer	Payment gateway unavailability, confirmation timeout, duplicate payment requests.	Loss of paid orders, risk of disputed transactions, increase in support requests.	Idempotent payment operations, failover between payment service providers, outbox pattern, asynchronous status reconciliation.
Inventory and logistics	Delayed synchronization with ERP/WMS, delivery API unavailability, incorrect product reservation.	Sale of unavailable goods, order cancellations, higher operational costs	Event-driven integration, eventual consistency, compensating operations, message queues
Promotions, recommendations and personalization	Failure of recommendation service, discount rules or external AI module.	Lower average order value and user engagement without complete sales interruption.	Graceful degradation, feature flags, static recommendations.
Admin panel and content management	CMS, PIM or back-office unavailability.	Delayed updates of assortment, prices and content.	Separation of front-office and back-office functions, publication queues, read-only mode.

The classification presented in table 1 shows that the architectural criticality of e-commerce components is determined not only by technical complexity, but also by their position in the purchasing journey. The cart, checkout and payment layer are the most vulnerable in terms of direct financial impact, as they convert user intent into a completed transaction. This is consistent with industry data: in 2026, Baymard Institute estimates the average cart abandonment rate at approximately 70%, indicating the high sensitivity of the checkout process to additional delays, errors or failures [3].

The practical importance of identifying critical failure zones is reinforced by the fact that digital service downtime has become a direct source of financial losses. According to the 2026 Splunk report published by Cisco, unplanned downtime costs Global 2000 companies an estimated $600 billion annually, while the average cost of downtime reaches approximately $15,000 per minute [4]. The same report states that 81% of organizations associated downtime incidents with customer loss. For e-commerce platforms, this risk is particularly significant, since users can quickly switch to competitors when the catalog, cart, checkout or payment functions are unavailable.

These data indicate that fault-tolerant e-commerce architecture should not be reduced to infrastructure redundancy alone. A more justified approach is to define an acceptable failure mode for each component in advance: full redundancy and rapid recovery for checkout and payment services, controlled degradation for recommendations and promotional mechanisms, and read-only or cached operation for catalog and content modules. Such differentiation makes it possible to preserve the purchasing journey even when individual services become partially unavailable.

Patterns for preventing cascading failures and system overload

A cascading failure in an e-commerce system occurs when the failure or slowdown of one component begins to overload related services: checkout waits for a response from the payment gateway, the cart is blocked by stock validation, search increases the load on the database, while repeated user requests and automatic retries further intensify system degradation [5]. Therefore, fault tolerance in backend architecture should be designed not only through redundancy, but also through mechanisms that limit error propagation. For this purpose, architectural patterns are used to interrupt unsuccessful calls, isolate resources, restrict incoming load and move part of the operations to an asynchronous mode (table 2).

Table 2

Architectural patterns for preventing cascading failures and system overload

Pattern / mechanism	Architectural task	Application in e-commerce	Expected effect
Circuit breaker	To stop repeated calls to a degraded service.	Payment gateway, search, delivery service, external APIs.	Prevention of cascading failures and release of threads, connections and other limited resources.
Bulkhead isolation	To isolate resources between functional domains.	Separation of checkout, catalog, search, recommendations and back-office tasks.	Localization of failure without stopping the entire commerce platform.
Timeout and retry with backoff	To handle short-term network or service failures without overloading dependencies.	Requests to ERP/WMS, PSP, delivery, pricing and inventory services.	Reduction of false failures without creating a retry storm.
Rate limiting and admission control	To limit incoming load to the level the system can process.	Peak sales, bot traffic, aggressive search requests, promotional campaigns.	Stabilization of latency and protection of checkout from non-critical traffic.
Backpressure	To signal overload upstream in the request or event-processing chain.	Order processing, inventory updates, email/SMS notifications.	Prevention of unprocessable queues and memory exhaustion.
Asynchronous queues and outbox pattern	To decouple the critical transaction from secondary integrations.	Order creation, notifications, ERP synchronization, analytics.	Preservation of the order event even when secondary systems are temporarily unavailable.
Graceful degradation	To preserve the purchasing journey when secondary services fail.	Recommendations, reviews, personalization, complex promotions and search facets.	Continuation of sales despite reduced completeness of user experience.

These patterns are particularly important for e-commerce platforms that combine a monolithic core, modular backend logic and multiple external services. In Magento 2 / Adobe Commerce architectures, critical dependencies include not only PHP modules for checkout and catalog management, but also OpenSearch, Redis, RabbitMQ, payment providers, ERP/WMS systems and delivery services. If these dependencies are called synchronously and without limits, even a single delay may exhaust worker pools, increase queue length, block user sessions and reduce conversion. Therefore, a more resilient approach requires each external call to have a defined timeout, a limited number of retries, a dedicated circuit breaker and a fallback scenario.

The relevance of these solutions is supported by current recommendations for designing distributed systems. In the **Microsoft Azure Architecture Center**, the circuit breaker pattern is described as a mechanism for temporarily blocking calls to a faulty remote service after errors are detected in order to prevent repeated unsuccessful requests and allow the system to recover [6]. For e-commerce, this means that the unavailability of a delivery or recommendation service should not automatically stop checkout if cached rates, a simplified ordering scenario, or delayed status synchronization can be applied.

The practical validation of such solutions should include not only standard load testing, but also testing system behavior under failure conditions. For example, the **Google SRE** guide emphasizes that overload is one of the main causes of cascading failures, while capacity planning should be combined with tests up to the breaking point and beyond-failure scenarios, since the specific mechanism of degradation is difficult to predict in advance [7]. For e-commerce, this means that testing should model not only “normal” high traffic, but also situations in which the payment gateway slows down, repeated requests increase, queues overflow and search performance degrades at the same time.

Thus, preventing cascading failures in e-commerce architecture requires a combination of several protection layers. Circuit breaker and timeout mechanisms limit the duration of impact from a failed dependency; bulkhead isolation prevents failure propagation across domains; rate limiting and backpressure protect the system from overload; queues and the outbox pattern make it possible to preserve a business event even when secondary systems are temporarily unavailable. This approach is consistent with the broader logic of architectural optimization in distributed systems, where improving the resilience of a digital service is considered not only a technical task, but also a factor in reducing operating costs and increasing economic efficiency [8]. Together, these mechanisms not only improve the technical stability of the platform, but also help preserve the key purchasing journey under partial infrastructure degradation.

Controlled degradation as a way to preserve sales under partial service unavailability

Controlled degradation in e-commerce architecture is an approach in which, under partial failure, the system preserves not the full range of functionality, but primarily the critical purchasing path [9]. From a technical perspective, this means the prior separation of functions into mandatory and secondary ones: the catalog, product page, cart, checkout and payment layer should remain operational for as long as possible,

whereas recommendations, reviews, complex promotional mechanisms, personalization, advanced filters and some analytics tracking may be temporarily disabled or replaced with simplified scenarios (fig. 1).

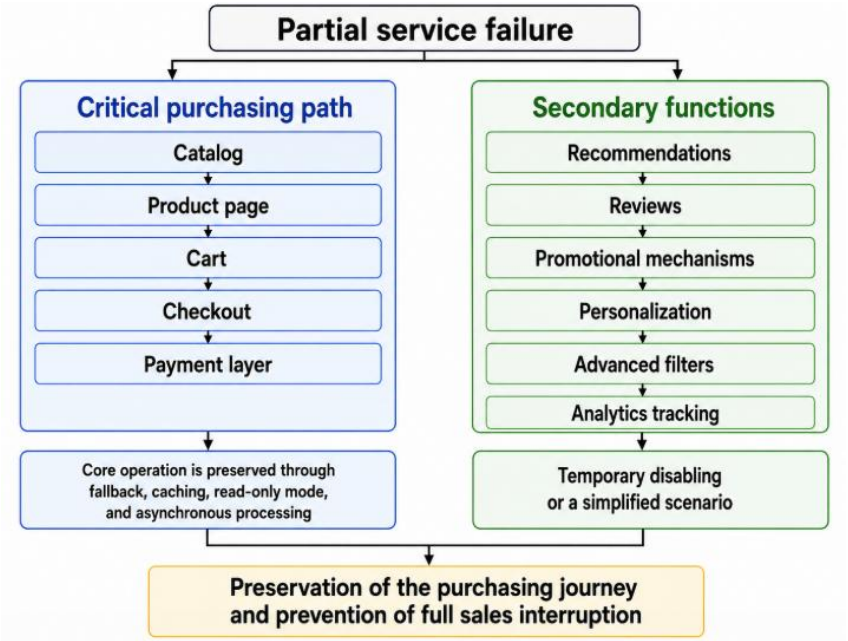


Figure 1. Controlled degradation in e-commerce architecture under partial service failure.

This logic is particularly important for platforms with a large number of external dependencies, where full synchronicity between services increases the risk of blocking a user session when one component fails.

In backend architecture, controlled degradation is implemented through fallback data, feature flags, read-only modes, caching, asynchronous processing, and the separation of critical and non-critical calls. For example, if the personal recommendation service is unavailable, the product page may display a static block of popular products; if an external delivery API fails, checkout may use cached rates or delayed validation; if the search engine degrades, a simplified output based on a basic index or a cache of popular queries may be applied. For platforms that use AI components in personalization, recommendations, or user scenario processing, it is especially important to define the boundaries of dependence on the intelligent service in advance, since the failure of such a module should not disrupt the execution of a core business operation [10]. In Magento 2 / Adobe Commerce-based systems, this is particularly relevant for modules that extend checkout, pricing, search, and integrations with ERP/WMS: such extensions should be designed so that their failure does not block order creation or the recording of a key business event.

Controlled degradation should be treated not as an emergency improvisation, but as a predefined operating mode of the platform. For each component, an acceptable failure behavior should be specified: idempotent reprocessing and asynchronous status reconciliation for payments; cached reading and limited update mode for the catalog; fallback to basic output for search; and temporary disabling of recommendations and personalization without affecting checkout. In a broader context, this approach is consistent with the use of information technologies to improve the operational efficiency of business processes, as it reduces dependence on manual recovery and on the full availability of all auxiliary services [11].

This approach also corresponds to user experience requirements. Google defines Core Web Vitals as metrics reflecting real user perception of loading, interactivity and visual stability, while in e-commerce even minor deterioration in speed or interactivity may affect conversion. Therefore, the architectural task is not only to recover after failure, but also to keep the purchasing journey sufficiently fast, predictable and complete when individual services are partially unavailable.

Recovery, observability and evaluation of the economic effect of fault tolerance

Recovery of an e-commerce platform after failure should be treated as a predesigned architectural process rather than a manual incident response. The system should define target recovery time, acceptable data loss, availability requirements for critical services, and user experience quality thresholds. At the backend level, this requires backup order-processing scenarios, idempotent operations, business event logging, message redelivery mechanisms, automatic failover, and procedures for restoring consistency between the catalog, cart, payments, inventory, and external integrations. In e-commerce, recovery must cover not only infrastructure components, but also the state of the commercial transaction: whether the order was created, the payment confirmed, the product reserved, and the status transferred to external systems.

A key condition for rapid recovery is **observability**, which integrates metrics, logs, traces, deployment events and business indicators into a unified view of platform state. For a commerce system, it is not enough to monitor CPU load, container availability or HTTP errors only; technical signals must be linked to business metrics such as checkout error rate, payment API response time, unfinished orders, inventory synchronization delay, queue depth, search failures and successful

transactions. According to the New Relic Retail & eCommerce Observability Forecast 2025, almost one in three retailers experiences critical outages weekly, while the median time to detect an outage in retail is 30 minutes and the median time to resolve it is 42 minutes. These figures show that, without end-to-end observability, a significant part of an incident may remain undetected before technical localization, which is especially risky during sales and peak demand periods.

The economic effect of fault tolerance should be assessed through reductions in detection time, recovery time and the number of affected user scenarios. In this context, observability becomes not only an operational tool, but also a mechanism for revenue protection. The same New Relic retail and e-commerce report states that 38% of retailers associate observability with improved revenue retention through deeper understanding of customer behavior, while 46% estimate the return on investment in observability at 2x or higher [12]. The report also notes tool consolidation: 51% of retailers reduce fragmented monitoring stacks, and the average number of observability tools decreased from 5.9 in 2022 to 3.9 in 2025. For e-commerce, this is important because fragmented monitoring complicates failure diagnosis across backend services, payment providers, search infrastructure, queues and external APIs.

Resilience assessment should also include engineering delivery metrics, since some failures are caused not by infrastructure incidents, but by changes introduced into the production environment. In e-commerce, this is especially important for releases affecting checkout, payment integrations, search indexes, pricing rules and database migrations. The main metrics used to evaluate release stability and recovery capability are summarized in table 3.

Table 3

Engineering delivery metrics for assessing e-commerce platform resilience

Metric	What it measures	E-commerce relevance
Failed deployment recovery time	Time required to restore service after a failed release.	Shows how quickly checkout, payment or search functionality can be recovered.
Change fail rate	Share of production changes that cause incidents or degradation.	Indicates how often releases disrupt the purchasing journey.
Deployment rework rate	Share of work spent on correcting released changes.	Reflects hidden engineering costs caused by unstable deployments.
Mean time to detect	Time between incident occurrence and detection.	Shows how quickly failures in critical services are identified.
Mean time to restore	Time required to return the system to normal operation.	Measures the effectiveness of recovery procedures.
Critical path availability	Availability of components required to complete an order.	Connects technical reliability with the preservation of sales.

The metrics presented in table 3 show that e-commerce resilience depends not only on infrastructure redundancy, but also on the stability of the release process. Therefore, fault-tolerant architecture should be supported by canary releases, blue-green deployment, automatic rollback, feature flags, API contract testing and database migration validation before production rollout.

Thus, the economic effect of fault tolerance is formed not only by preventing complete downtime, but also by reducing incident scope, accelerating diagnostics, decreasing manual engineering effort, and preserving transactional integrity. In practice, this requires linking technical and business metrics: SLOs for checkout and payments, error budgets for critical APIs, RTO/RPO for databases and queues, MTTA/MTTR for incident response, as well as indicators of unfinished orders and successful payments. In this model, fault tolerance becomes a measurable management parameter of an e-commerce platform, making it possible to justify investments in redundancy, observability, recovery automation and safe release practices through lower operational risks and more predictable digital sales.

Conclusion

Designing fault-tolerant e-commerce systems requires a comprehensive approach in which platform architecture is considered not only as technical infrastructure, but also as a mechanism for maintaining commercial continuity. The key elements of such architecture include identification of critical failure zones, prevention of cascading failures, dependency isolation, controlled degradation of secondary functions, end-to-end observability and predefined recovery scenarios. For a digital commerce platform, it is essential that the partial unavailability of search, recommendations, external APIs or back-office components does not stop the catalog, cart, checkout and payment layer.

Therefore, backend engineering solutions, including circuit breaker, message queues, operation idempotency, feature flags, caching, failover and safe deployment practices, should be evaluated through their impact on preserving the purchasing journey, reducing operational risks and preventing revenue losses. In this context, fault tolerance is not an auxiliary system property, but one of the key factors in the sustainable development of an e-commerce business and in strengthening user trust in the digital commerce platform.

References

1. Berezhnoy A. Architectural approaches to designing scalable information systems in e-commerce. *International Journal of Advanced Trends in Computer Science and Engineering*. 2025;14(6):222-226. DOI: 10.30534/ijatcse/2025/011462025 EDN: YHONII.

2. Bathina S. Composable commerce architectures: Building agile retail systems. *International Journal of Advanced Research in Computer Science & Technology (IJARCST)*. 2025;8(3):12226-12231.

3. Cart & Checkout Usability Research. *Baymard Institute*. URL: https://baymard.com/research/checkout-usability (Accessed May 15, 2026).

4. The Hidden Costs of Downtime 2026. *Splunk*. URL: https://www.splunk.com/en_us/form/the-hidden-costs-of-downtime.html (Accessed May 15, 2026).

5. Zeng Z., Wang N., Xu D., Chen R. Cascading failure modeling and resilience analysis of coupled centralized supply chain networks under hybrid loads. *Systems*. 2025;13(9):729. DOI: 10.3390/systems13090729 EDN: LZJFXS.

6. Circuit Breaker pattern. *Microsoft Azure Architecture Center*. URL: https://learn.microsoft.com/en-us/azure/architecture/patterns/circuit-breaker (Accessed May 16, 2026).

7. Addressing Cascading Failures. *Google SRE*. URL: https://sre.google/sre-book/addressing-cascading-failures/ (Accessed May 16, 2026).

8. Kovalevskyi K. Impact of architectural optimization of distributed systems on reducing operating costs and improving the economic efficiency of digital services. *International Journal of Engineering in Computer Science*. 2026;8(2):63-67.

9. Miranda F. D. S., dos Santos D. S., Vilela R. F., Assunção W. K. G., dos Santos R. C., Pinto V. H. S. C. A proposed catalog of development patterns for fault-tolerant microservices. Proceedings of the XXIII Brazilian Symposium on Software Quality. 2024:406-416.

10. Miloserdov A. Methodological foundations of designing AI platforms for corporate B2B users. *Professional Bulletin: Information Technology and Security*. 2026;(1):43-52.

11. Lim E. Use of information technologies to improve the operational efficiency of business processes in the financial sector. *Universum: Economics and Jurisprudence: electronic scientific journal*. 2026;5(139):52-60.

12. New Relic Report Reveals Retailers are Turning to Observability Amidst AI Adoption. *New Relic*. URL: https://newrelic.com/press-release/20251104 (Accessed May 18, 2026).

No 189 (189) (2026)

The scientific heritage
(Budapest, Hungary)
The journal is registered and published in Hungary.
The journal publishes scientific studies, reports and reports about achievements in different scientific fields.
Journal is published in English, Hungarian, Polish, Russian, Ukrainian, German and French.
Articles are accepted each month.
Frequency: 24 issues per year.
Format - A4
ISSN 9215 — 0365

All articles are reviewed
Free access to the electronic version of journal
Edition of journal does not carry responsibility for the materials published in a journal.
Sending the article to the editorial the author confirms it's uniqueness and takes full responsibility for possible consequences for breaking copyright laws

Chief editor: Biro Krisztian
Managing editor: Khavash Bernat

- Singula Aleksandra - Professor, Department of Organization and Management at the University of Zagreb (Zagreb, Croatia)
- Bogdanov Dmitrij - Ph.D., candidate of pedagogical sciences, managing the laboratory (Kiev, Ukraine)
- Chukurov Valeriy - Doctor of Biological Sciences, Head of the Department of Biochemistry of the Faculty of Physics, Mathematics and Natural Sciences (Minsk, Republic of Belarus)
- Torok Dezso - Doctor of Chemistry, professor, Head of the Department of Organic Chemistry (Budapest, Hungary)
- Filipiak Pawel - doctor of political sciences, pro-rector on a management by a property complex and to the public relations (Gdansk, Poland)
- Flater Karl - Doctor of legal sciences, managing the department of theory and history of the state and legal (Koln, Germany)
- Bence Orban - Doctor of sociological sciences, professor of department of philosophy of religion and religious studies (Miskolc, Hungary)
- Feld Ella - Doctor of historical sciences, managing the department of historical informatics, scientific leader of Center of economic history historical faculty (Dresden, Germany)
- Owczarek Zbigniew - Doctor of philological sciences (Warsaw, Poland)
- Gál Jenő - MD, assistant professor of history of medicine and the social sciences and humanities (Budapest, Hungary)
- Borbély Kinga - Ph.D, Professor, Department of Philosophy and History (Kosice, Slovakia)
- Eberhardt Mona - Doctor of Psychology, Professor, Chair of General Psychology and Pedagogy (Munich, Germany)
- Kramarchuk Vyacheslav - Doctor of Pharmacy, Department of Clinical Pharmacy and Clinical Pharmacology (Vinnytsia, Ukraine)

«The scientific heritage»
Editorial board address: Budapest, Kossuth Lajos utca 84,1204
E-mail: public@tsh-journal.com
Web: www.tsh-journal.com